The derivation of a valence forcefield for carbohydrates

Viner, Russell

Award date:
1989

Awarding institution:
University of Bath

THE DERIVATION OF A

VALENCE FORCEFIELD

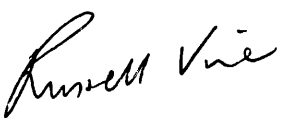
FOR CARBOHYDRATES

submitted by ***RUSSELL VINER***

for the degree of Ph.D.

at the University of Bath

1989



(Russell Viner)

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

The work presented in this thesis was undertaken at the University of Bath between August 1986 and July 1989. Funding from the Science and Engineering Research Council is gratefully acknowledged.

I must express my gratitude to all those who have assisted me in the course of this work. I would like to thank my supervisor, David Osguthorpe, for his ideas and guidance. I am particularly indebted to Pnina Osguthorpe, who has been an invaluable source of advice and encouragement. She also took on the onerous task of proof-reading this thesis. My colleagues in the Molecular Modelling Unit also deserve a mention - Prem Paul, Richard Sessions, Paul Burney, Christina Hennecke and Vic Cockcroft - for making my three years there so enjoyable.

Finally, I would also like to thank both Nina and my parents, for their continued encouragement and support.

Abstract

The Derivation of a Valence Forcefield for Carbohydrates

A new forcefield has been developed for modelling the conformational and dynamic behaviour of carbohydrates. The *anomeric* and *gauche* effects (present in compounds containing geminally and vicinally di-substituted electronegative atoms) are important in determining the conformations of carbohydrate molecules and these are accounted for in the new forcefield. In particular, the anomeric effect is represented in the forcefield function by a new *bond-torsion* cross term. This is demonstrated to reproduce both the relative energies as well as the changes in bond lengths exhibited by the various rotameric forms of compounds containing an anomeric centre.

The forcefield parameters have been systematically fitted to the experimental data of a large range of model compounds consisting of hydrocarbons, ethers, acetals and alcohols that contain the structural features found in carbohydrate molecules.

The database of observables used in deriving the forcefield was selected to reflect not only the static properties associated with equilibrium structures but also those concerned with molecular motion (e.g. vibrational frequencies and rotational barriers).

Molecular geometries determined by gas phase electron diffraction are shown to be reproduced well by the forcefield. Calculated frequencies have been extensively fitted to the vibrational spectra of small symmetrical molecules for which the assignment of the vibrations is less ambiguous due to symmetry considerations. Rotational barriers and conformational energy differences calculated by the forcefield are shown to agree with experimental values.

Thermodynamic and structural crystal properties such as heats of sublimation and unit cell parameters are also calculated and demonstrate good agreement with those observed experimentally. These are a test of the suitability of the forcefield for modelling intermolecular interactions and have often been overlooked in previous forcefields.

The new forcefield thus gives a good account of both the structural and dynamic features of carbohydrate molecules and should prove a useful tool in the conformational analysis of this class of compounds.

Contents

Chapter 1

Introduction

1.1 Why Model Carbohydrates ?

Until a few years ago, carbohydrates were regarded as being only important for energy storage and metabolism. However, they are increasingly becoming seen as the equal partners of proteins and nucleic acids in terms of biological importance.[1]

In recent years, carbohydrate molecules have been identified as information carriers and recognition molecules in many areas of biochemistry. Many proteins for instance, exist in the body not as naked proteins but rather as *glyco-conjugates*, bearing carbohydrate side-chains that are often essential to the biological activity. In some cases, the proteins themselves act merely as platforms for the glycosidic chains. The carbohydrates responsible for blood group specificity, for example, if attached to a synthetic polymer rather than the native protein, will still evoke the same blood group antigenicity.[1]

Carbohydrates have also been found to have regulatory functions in organisms as diverse as plants, fungi and bacteria. A class of oligosaccharide plant hormones, the *oligosaccharins*, have been identified by Albersheim to be of central importance to the growth, development and reproduction of plants, as well as in defence against disease.[2]

Other functions that have been attributed to carbohydrates are as receptors for binding toxins, viruses and hormones. They are also known to alter drug pharmacokinetics, control vital events in fertilisation, and target aging cells for destruction.[3]

The majority of carbohydrates are composed of only a handful of saccharide residues: glucose, mannose, galactose and so forth. Even though the composition of carbohydrates are so similar, they are seen to display a wide diversity in their biological functions. As with proteins, this diversity of function comes from the structure and conformation of these molecules, and it is their three-dimensional shapes that govern their biological activity. This has been appreciated for a long time, and a very great deal of effort has been applied to the experimental elucidation of the conformations of biopolymers.

The usefulness of theoretical models for biopolymer structures was also recognised at an early stage.[4] Since then, extensive use of molecular modelling has been made in the study of protein structure in particular.[5-8] As a result, the theoretical models of proteins have reached a high level of refinement.[9-11]

Carbohydrates have not received the same level of interest as proteins or nucleic acids either theoretically or experimentally; and this prompted Goodall and Norton to describe them in a recent paper as the 'Cinderella' of the biopolymers.[12] However, with the importance of carbohydrates in all areas of biology becoming increasingly evident, interest in carbohydrate conformation will also continue to grow, and it is therefore desirable that reliable methods for modelling them should be developed.

1.2 Molecular Modelling Methods

Generally speaking, all molecular modelling methods attempt to describe the properties of a molecular system in terms of a mathematical function of the atomic positions. Such a theoretical model, if sufficiently accurate, can then be used to *predict* a variety of information about the molecular system. For example, various molecular modelling techniques can been used to calculate the minimum energy conformation of a molecule, its molecular geometry, relative energies of different conformations, molecular

dipole moments and even its vibrational spectrum.

The methods used in molecular modelling can be separated into two broad categories:

(i) Quantum Mechanical Calculations

(ii) Empirical Energy Calculations

These categories differ in the way in which the mathematical models describing the molecular system are derived. The first category, quantum mechanical calculations, approach the problem from a purely theoretical standpoint. These methods attempt to apply the principles of quantum mechanics to define the mathematical model of the molecular system. Because of the complexities involved in applying quantum mechanics to all but the simplest molecules, a variety of approximations and simplifications are often made to make the calculations more tractable.[13]

The second category, as the word 'empirical' suggests, includes methods that are not derived on purely theoretical grounds, but rather by selecting a mathematical model that from empirical considerations should give a reasonable representation of molecular behaviour. This model is then refined by fitting it to known experimental data. The assumption made is that if the model can be made to reproduce a range of known experimental data with reasonable accuracy, it can then be used to predict similar, but as yet unmeasured properties, with an equivalent degree of accuracy.

There is therefore a wide diversity in molecular modelling methods, varying from relatively simple empirical methods that can be applied even to large molecules such as biopolymers, to complex quantum mechanical calculations requiring hours of computer time even for small molecules. Most of these methods have at some time or another been used in the study of carbohydrates (or model compounds for carbohydrates) and it is therefore worthwhile giving a brief appraisal of them.

1.2.1 Quantum Mechanical Methods

Although this thesis is concerned primarily with an empirical modelling method, quantum mechanics has made important contributions to the study of carbohydrates; particularly in unravelling the mechanisms behind the *anomeric* and *gauche* effects (discussed in Chapters 6 & 7). In addition, comparisons between our forcefield calculations and those of quantum mechanical calculations will often be drawn, and some understanding of them is therefore necessary.

In principle, of the various molecular modelling methods, quantum mechanical calculations are the most appropriate from theoretical considerations. According to quantum mechanics, the energy (E) of a stationary molecule may be obtained by a solution of the Schroedinger partial differential equation

$$\hat{H}\Psi = E\Psi$$

where $\hat{H}$ is the Hamiltonian, a differential operator representing the the total energy; and Ψ is the wavefunction of the molecule, and is dependent on the molecular geometry.

Quantum mechanical calculations generally use a method which assumes that the electronic components of the wavefunction (the molecular orbitals, ψ_i) are combinations of the atomic orbital wavefunctions, ϕ_j. Thus, for a given molecular orbital, ψ_i

$$\psi_i = c_1\phi_1 + c_2\phi_2 + c_3\phi_3 + \cdots c_j\phi_j + \cdots c_n\phi_n$$

where the value denoted c_j is the *coefficient* of atomic orbital ϕ_j in the molecular orbital ψ_i. This method is known as the *linear combination of atomic orbitals (LCAO)* approximation. Given this approximation, the best set of coefficients can be determined using the *Variation Theorem*, which states that the set of coefficients for all the

molecular orbital wavefunctions (ψ_i) will be the one that gives the lowest total energy E.

Calculations of this sort are known as *ab initio* calculations because they construct a model of the electronic nature of the molecule from 'first principles', and do not rely on any experimentally derived knowledge.

The atomic orbitals (ϕ) are described by a set of mathematical functions called the *basis set*. These mathematical functions vary in their level of complexity, and in general, more complex functional forms will give a more accurate description of the atomic orbitals, but at an increased cost in computer time. Several references to *ab initio* calculations are made in this text, and the basis set used in each case is specified. Common standard basis sets used, in order of increasing sophistication, are STO-3G, 4-21G and 6-31G. The 6-31G basis set, with its various modifications, represents the most sophisticated level of calculations in general use. (For a full discussion of *ab initio* basis sets, see reference 13).

An advantage of *ab initio* calculations is that they give a complete description of the electronic nature of the molecule, and therefore a wider range of molecular properties can be deduced than from empirical calculations. In addition, *ab initio* calculations, unlike empirical methods, do not have to be fitted to experimental data, and can therefore be directly applied to molecular systems for which no experimental information is known. The major drawback, however, is the computational expense. Although a full *ab initio* geometry optimisation of, say, a pyranose ring is today feasible, it would take hours of supercomputer time. A full conformational analysis for such a molecule becomes unquestionably too large, and (for the time being) empirical energy calculations must be used if the conformational analysis of large molecules is to be undertaken.

1.2.2 Empirical Energy Calculations

Rigid Geometry Calculations

The concepts of steric repulsion, electrostatic interactions and the preference for staggered rotations about single bonds have long been used by chemists to rationalise (and predict) the conformational preferences of molecules. These concepts, at first qualitative, were eventually quantified so that conformational energies could be calculated for a given molecular geometry. Thus the total conformational energy for a molecule can be estimated from the following sum:

$$E_{Total} = E_{VDW} + E_{Elec} + E_{Torsion}$$

where

$$E_{VDW} = \Sigma \text{ (atom-atom van der Waals interactions)}$$

$$E_{Elec} = \Sigma \text{ (atom-atom electrostatic interactions)}$$

$$E_{Torsion} = \Sigma \text{ (torsional energies)}$$

The precise mathematical functions describing the van der Waals interactions, the electrostatic interactions and the torsional energies vary from one method to another, but are generally fitted to known experimental data on conformational energies.

Methods of this type have become known as *Rigid Geometry* calculations, because they assume that bond lengths and valence angles remain constant regardless of conformation. This assumption is made on the basis that bond lengths and valence angles are 'stiff' in comparison with torsion angles, and will not be distorted much by conformational changes. Rigid Geometry methods therefore neglect any contribution to the conformational energy that may actually occur in the molecule due to distortions of this type. The principal advantage of Rigid Geometry calculations are that they contain very few energy terms that need to be calculated, and are therefore

reasonably fast and may be applied to large molecular systems.

Flexible Geometry Calculations

With the advent of faster and more powerful computers in recent years, it has become feasible to use more complex energy functions than those used in Rigid Geometry, and so avoid some of the larger approximations made in that method. Flexible Geometry calculations extend the Rigid Geometry method by adding energy terms that take account of the distortion of bond lengths and valence angles:

$$E_{Total} = E_{VDW} + E_{Elec} + E_{Torsion} + E_{Bond} + E_{Angle}$$

where

$$E_{Bond} = \Sigma \text{ (Bond stretch energies)}$$

$$E_{Angle} = \Sigma \text{ (Angle bend energies)}$$

More elaborate energy functions also include *cross terms*, energy terms that allow for the fact that distortions in the internal coordinates of a molecule are not independent of each other. Thus the stiffness of a given bond will be to some extent dependent on the distortion in adjacent bonds, angles and so forth.

In addition to their relationship to Rigid Geometry, Flexible Geometry energy functions also owe part of their ancestry to the *valence forcefield* equations that have been used for many years in the vibrational analysis of molecules.[14,15] These force-field equations are similar to those of Flexible Geometry but for the absence of the non-bond terms, E_{VDW} and E_{Elec}. It is because of this relationship that the terms *forcefield* and *valence forcefield* are often used to describe Flexible Geometry energy functions. [Note: The term *molecular mechanics* has also been gaining popularity in recent years to describe the Flexible Geometry method.[16] However, we shall use the

terms *forcefield* and *forcefield calculations* in this text.]

Flexible Geometry forcefield calculations have increasingly become the preferred molecular modelling tool for the study of biopolymers such as proteins[8] and nucleic acids.[10] Biopolymers in some ways lend themselves to the application of forcefield calculations because they are composed of similar units: amino acids in the case of proteins, and nucleotides in the case of nucleic acids. This limits the number of atom types and functional groups that forcefield parameters must be found for. Carbohydrates are also good candidates for forcefield calculations: they generally consist of only three elements, carbon, hydrogen and oxygen, in the form of pyranose and furanose rings. The practical problem of forcefield calculations is one of reliability. A forcefield is developed by fitting the equations and parameters to reproduce experimental data. As long as the forcefield is then applied to molecules similar to those from which the experimental data was obtained, the results should be of comparable accuracy to those of the original fit. On the other hand, if we attempt to extrapolate the forcefield for use beyond the areas for which it has been tested, the reliability of the calculations will be in doubt.

The development of a forcefield by fitting to experimental data is a very labour intensive task, as the forcefield must be repeatedly revised until an acceptable fit to the data is achieved. This thesis documents such a task; the development of a forcefield for carbohydrates.

An increasing application of forcefields is in *molecular dynamics*. In molecular dynamics, the forces on each atom of a system are calculated using the forcefield. Using these forces, Newton's equations of motion may then be solved to give a description of the dynamic behaviour of the molecular system.[8] Such simulations may be applied to many problems which involve determining the accessible conformations of molecules, including biopolymers. Dynamic simulations are becoming an increasingly important tool in the design of peptides, drugs and other biologically active

molecules.[17]

1.3 Previous Molecular Modelling of Carbohydrates

Rather than giving an extensive review of carbohydrate modelling, we shall instead just highlight a few of the applications of the methods described above.

Rigid Geometry studies of α-D-glucose by Ramachandran's group were amongst the first computational studies made on carbohydrate systems.[18] This study indicated the $^4C_1$ conformation (i.e with all ring substituents equatorial except for the anomeric hydroxyl) to be the most stable. It also encountered one of the main problems in the calculation of carbohydrate conformations; that is, the very large number of possible conformations resulting from rotation about the hydroxyl groups. This is a manifestation of the *multiple minima* problem: highly flexible molecules often have a multitude of local minima that make finding the global minimum difficult.

Another type of Rigid Geometry calculation that has been used fairly widely is the HSEA method of Lemieux and co-workers. This has been applied to the conformational analysis of oligosaccharides, in particular those associated with blood group determination.[19-21] It also endeavours to account for the conformational preferences of saccharide chains caused by the anomeric effect, and this is discussed in more detail in Chapter 7.

Flexible Geometry forcefield calculations have also been made for a number of carbohydrate systems. Most notable perhaps, is the forcefield developed by Rasmussen.[22] This is the only forcefield to have been developed specifically for carbohydrates, and is certainly one of the best documented.[23] The functional form of the energy function is a fairly simple one, but nevertheless it gives a good representation of molecular geometry in particular, as was its principal intention. The Rasmussen forcefield has been applied to a variety of monosaccharide[24] and disaccharide systems,[25-27] and Brady has also adapted it to perform molecular dynamics

calculations.[28]

Perhaps the best known forcefields are those of Allinger, MM1 and MM2. However, these have more often been applied to small organic molecules rather than biopolymers.[16] MM1 was parameterised for use with alcohols and ethers,[29] and the resulting forcefield applied to pyranosides by Jeffrey and Taylor.[30] This gave satisfactory results, with the exception of the geometry at the anomeric carbon atoms. A modification was made in order to overcome this, but at the expense of a large number of additional bond length and angle parameters.[30] Allinger and Norskov-Lauritsen have also modified the MM2 forcefield to account for the anomeric effect, and this is compared with our own representation of the anomeric effect in Chapter 7. Unfortunately, the MM1 and MM2 forcefields cannot easily be used to perform molecular dynamics calculations. This is because they treat lone-pair electrons as pseudo-atoms with zero mass, and the molecular mechanics algorithm uses the mass of the atoms to determine their acceleration. A mass of zero results in an infinite acceleration, and so the new position of the lone-pair cannot be determined.

Although most biologically important carbohydrates are too large to be handled by quantum mechanical calculations; *ab initio* methods have nevertheless contributed to our understanding of carbohydrate conformation through the study of smaller, model compounds. In particular, Pople, Radom and Hehre have performed extensive calculations in order to determine the nature of the anomeric effect.[31-33] The anomeric effect is a stereoelectronic effect that affects the conformation about the anomeric carbon atoms in pyranose and furanose rings.[34]

1.4 Objectives of this Study

At Bath University, the molecular modelling group currently employs a forcefield known as the VFF ('valence forcefield') in the study of protein and peptide conformation.[11] This forcefield was developed and parameterised for model compounds

exhibiting the structural features found in amino acids,[35-38] and has subsequently been successfully employed in the study of large protein molecules,[11,39] as well as in the molecular dynamics of peptide hormones.[40]

For the reasons discussed in Section 1.1, carbohydrates have also become an appealing candidate for molecular modelling by forcefield calculations. The objectives of this study were therefore to extend the VFF forcefield for use with carbohydrates, in order that their structure, energetics and dynamical properties can be simulated with the same reliability as is currently possible for protein and peptide systems.

This work has entailed the fitting of the forcefield to a data base of model compounds for carbohydrate molecules; consisting of hydrocarbons, ethers, alcohols and acetals. Because certain stereoelectronic effects are central to the conformational behaviour of carbohydrates (the anomeric and *gauche* effects) they represent something of a challenge for forcefield calculations. It was recognised at the outset of this project that new functional forms for the forcefield might have to be developed in order to accommodate them.

1.5 References to Chapter 1

1. P. Knight, *Bio/Technology*, **7(1)**, 35 (1989).
2. P. Albersheim and A.G. Darvill, *Scient. Am.*, 44 (Sept. 1985).
3. T.W. Rademacher, R.B. Parekh, and R.A. Dwek, *Glycobiology, Ann. Rev. Biochem.* , **57**, 785 (1988).
4. G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, **7**, 95 (1963).
5. H.A. Scheraga, in *Current Topics in Biochemistry 1973*, ed. C.B. Anfinsen & A.N. Schechter, p. 1, Academic, New York (1974).

6. B. Robson and D.J. Osguthorpe, *Amino Acids, Pept. Proteins*, **9**, 196 (1978).

7. A.J. Hopfinger, in *Conformational Properties Of Macromolecules*, Academic, New York (1973).

8. A.T. Hagler, *Theoretical Simulation of Conformations, Energetics and Dynamics of Peptides,* in *The Peptides*, vol. 7, p. 213, Academic, New York (1985).

9. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).

10. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).

11. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).

12. D.M. Goodall and I.T. Norton, *Acc. Chem. Res.*, **20**, 59 (1987).

13. W.J. Hehre, L. Radom, P.v.R. Schleyer, and J.A. Pople, in *Ab Initio Molecular Orbital Theory*, Wiley-Interscience (1986).

14. P. Gans, in *Vibrating Molecules*, Chapman & Hall, London (1971).

15. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations,* McGraw Hill, New York (1955).

16. U. Burkert and N.L. Allinger, in *Molecular Mechanics, ACS Monograph 177*, American Chemical Society, Washington, D.C. (1982).

17. D.J. Osguthorpe, P. Dauber-Osguthorpe, R.B. Sessions, P.K.C. Paul, and P.A. Burney, in *Topics in Medicinal Chemistry, 4th SCI-RSC Medicnal Chemistry Symposium*, ed. P.R. Leeming, p. 332, R.S.C., London (1988).

18. V.S.R. Rao, P.R. Sundararajan, C. Ramakrishnan, and G.N. Ramachandran, in *Conformation of Biopolymers*, p. 721, Academic, London (1967).

19. R.U. Lemieux and S. Koto, *Tetrahedron*, **30**, 1933 (1974).

20. H. Thogersen, R.U. Lemieux, K. Bock, and B. Meyer, *Can. J. Chem.*, **60**, 44 (1982).

21. R.U. Lemieux and K. Bock, *Archiv. Biochem. Biophys.*, **221**, 125 (1983).

22. S. Melberg and K. Rasmussen, *J. Mol. Struc.*, **57**, 215 (1979).

23. S. R. Niketic and K. Rasmussen, in *The Consistent Force Field*, Springer, New York (1977).

24. K. Rasmussen, *Acta Chem. Scand.*, **A33**, 323 (1980).

25. S. Melberg and K. Rasmussen, *Carbohydr. Res.*, **69**, 27 (1979).

26. S. Melberg and K. Rasmussen, *Carbohydr. Res.*, **71**, 25 (1979).

27. S. Melberg and K. Rasmussen, *Carbohydr. Res.*, **78**, 215 (1980).

28. J.W. Brady, *J. Am. Chem. Soc.*, **108**, 8153 (1986).

29. N. L. Allinger, S. H.-M. Chang, D. Hindman Glaser, and Helmut Honig, *Israel J. Chem.*, **20**, 51-56 (1980).

30. G.A. Jeffrey and R. Taylor, *J. Comput. Chem.*, **1**, 99 (1980).

31. L. Radom, W.J. Hehre, and J.A. Pople, *J. Am. Chem. Soc.*, **94**, 2371 (1972).

32. G.A. Jeffrey and J.H. Yates, *J. Am. Chem. Soc.*, **101**, 820 (1979).

33. G.A. Jeffrey , J.A. Pople, and L. Radom, *Carbohydr. Res.*, **25**, 117 (1972).

34. A.J. Kirby, in *The Anomeric Effect and Related Stereoelctronic Effects at Oxygen*, Springer, Berlin (1983).

35. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).

36. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979).

37. A. T. Hagler, S. Lifson, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5122 (1979).

38. A. T. Hagler, P. Dauber, and S. Lifson, *J. Am. Chem. Soc.*, **101**, 5131 (1979).

39. M.M. Campbell, J. Long-Fox, D.J. Osguthorpe, M. Sainsbury, and R.B. Sessions, *J. Chem. Soc., Chem. Commun.*, 1560 (1988).

40. B.I. Baker, D.W. Brown, M.M. Campbell, R.G. Kinsman, C.A. Moss, D.J. Osguthorpe, P.K.C. Paul, and P.D. White, *J. Chem. Soc., Chem. Commun.*, 1543 (1988).

Chapter 2

Strategy for Developing the Forcefield

2.1 Requirements of the Model

Before attempting to construct any theoretical model, it is important that the purpose of the model be fully defined. In order to derive a forcefield, we must therefore have a clear idea of the uses it will be put to, and the type of information it will be expected to provide.

As with the protein and nucleic acid forcefields already in use,[1-3] the purpose of a carbohydrate forcefield will be to simulate conformational behaviour. The term 'conformational behaviour' covers a broad area, but is used here to represent the following four areas:

(i) Molecular Structure

(ii) Conformational Energies

(iii) Molecular Motion

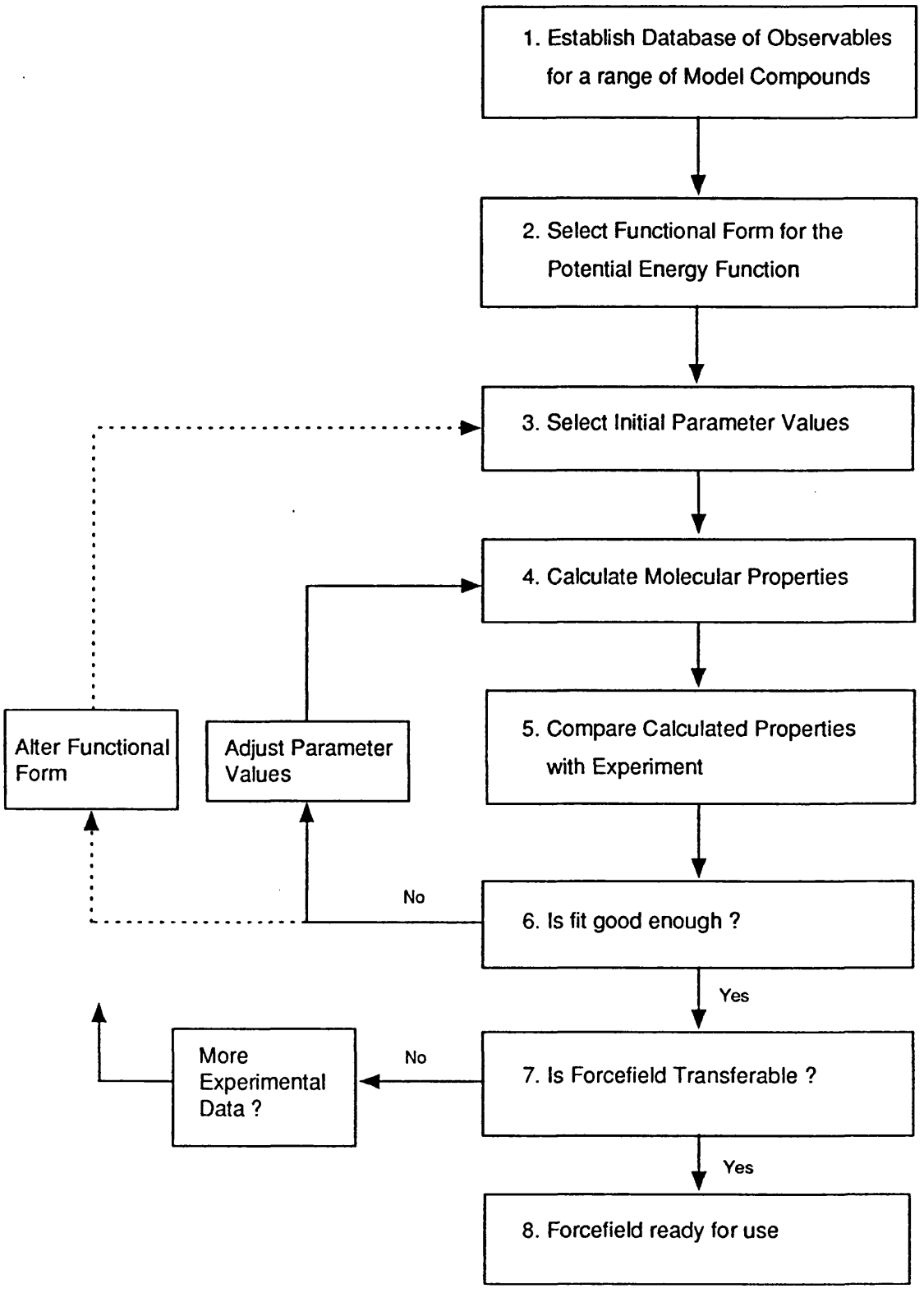
(iv) Intermolecular Interactions

Each of these areas must be reproduced by the forcefield if it is to be useful in the study of the conformation, and hence the function, of carbohydrates.

2.2 Developing the Forcefield

The basic stages involved in the development of a forcefield are shown diagrammatically in Figure 2.1 (overleaf).

Figure 2.1 Steps Involved in the Development of a Forcefield



Establishing a Database of Observables

As with any empirically derived model, the first step in the development of a forcefield must be to establish a 'database of observables', that is, a collection of data to be used in fitting the forcefield. In Chapter 1 it was emphasised that empirically derived models are only reliable when predicting similar properties to the ones that they were originally fitted to. The experimental data to which we fit our forcefield should therefore correspond as closely as possible to the type of information we will want it to provide, when applied predictively to new systems. (Setting up the database of observables will be discussed at greater length in the latter half of this chapter.)

Selection of the Functional Form

One of the main assumptions of forcefield calculations is that the conformational energy of a molecule may be partitioned into a set of energy terms that each have some physical significance. In the first chapter, a typical potential energy function was given as:

$$E_{Total} = E_{VDW} + E_{Elec} + E_{Torsion} + E_{Bond} + E_{Angle}$$

Each of these energy components is related in some way to the geometry of the molecule. For example, the van der Waals energy (E_{VDW}) is generally regarded as some function of the interatomic distances between non-bonded atoms within the molecule; but the precise form of this function must be decided on. These decisions are for the most part based on theoretical considerations; and in the case of the van der Waals energy, our forcefield regards it to be the sum of Lennard-Jones terms between atom pairs:

$$E_{VDW} = \sum \varepsilon_{ij} \left[\left(\frac{r^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^*_{ij}}{r_{ij}} \right)^{6} \right]$$

The quantity r_{ij} in the above equation is the interatomic distance between a pair of atoms i and j. r^*_{ij} and ε_{ij} are examples of the forcefield *parameters* - constants whose values have to be carefully chosen to fit experimental data.

Practical considerations can also have a bearing on the choice of functional form, however, as they should not be so complex that too much computer time will be taken in calculating them. This is particularly important for the non-bond terms (van der Waals and electrostatic) as they generally form the bulk of the calculation. The number of interatomic non-bond interactions is roughly proportional to the number of atoms squared. In large molecular systems, a great many such interactions therefore need to be computed, and so non-bond energy terms should ideally be as simple as possible while still maintaining accuracy.

In the present study, we were extending an existing forcefield for use with a new class of compounds, and therefore the functional form was for the most part already determined.[3] Some additions have had to be made to accommodate carbohydrates (see Chapter 7) and the full functional form of the potential energy function is given in Chapter 3 (Equation 3.1).

Selection of the Initial Parameter Values

Having decided on an appropriate functional form, initial parameter values need to be selected. Starting values can be estimated from various sources, including vibrational spectra, structural data or rotational barriers.[4] Another source of initial parameter values is from other closely related forcefields: in the present work, initial parameter values were taken mostly from the VFF forcefield[3] (see Chapter 4).

Optimisation of the Forcefield

Once initial estimates for the parameters have been selected, the process of optimisation can begin. This consists of minimising the difference between the calculated properties and the experimental ones, and is performed by an iterative process, indicated by the loop in Figure 2.1. The first step in this iteration is the calculation of the properties for the selected model compounds. (The methods by which the various properties are calculated is the subject of Chapter 3.) Having calculated these properties, they can then be compared with the experimental values, and the differences between them (the *deviations*) determined. The parameter values are then adjusted in order to reduce these deviations, and the process repeated until an acceptable fit to the experimental data is achieved.

This iteration process may be either by trial-and-error, or by systematic optimisation using a least-squares procedure.[4] A combination of both methods was used in the development of our forcefield, as it is almost impossible to fully automate the optimisation of parameters. This is because certain decisions can only be made with the benefit of intuition that are difficult to include in an optimisation algorithm. The least-squares method is useful when a large number of parameters are being fitted to a large amount of data simultaneously, as is often the case when fitting vibrational frequencies. A fuller description of the optimisation of the parameters; and a description of the least-squares procedure, is made in Chapter 4.

Occasionally, optimisation of the parameter values may not be sufficient to fit the experimental data. This occurs when the functional form of the potential energy function is insufficient - in other words, the model is not a good representation of reality. In this event, the functional form must be reconsidered and altered accordingly (see the dashed line in Figure 2.1).

Transferability of the Forcefield

Once the parameter values have been optimised, the forcefield should then be tested for *transferability.* A forcefield is considered to be transferable across a range of molecules if the calculated properties for the molecules are all in similar agreement to experiment.[4] The assumption that such a transferable forcefield may be found is implicit in all empirical molecular modelling methods. Forcefield parameters are invariably derived by fitting to experimental data for small model compounds. This assumes that the structural features present in a small molecule will behave similarly to the same structural features in a larger molecule. For example, common model compounds for the **C–H** bond are the simple alkanes (methane, ethane, propane etc.)[5,6] and parameters for the **C–H** bond are derived from experimental data on these compounds. If the resulting forcefield is then applied to carbohydrates, the assumption is made that these **C–H** bond parameters are sufficient to model **C–H** bonds within carbohydrates. In practice, this assumption is seen to be a reasonable one, as bond lengths, vibrational frequencies and so forth, are very similar for **C–H** bonds in both alkanes and carbohydrates.

Occasionally, the parameters for a particular structural unit will not be sufficiently transferable. This most often occurs when the environment of the structural unit changes significantly from one molecule to another. In the present work, for instance, the **C–O** bond in acetals could not be well reproduced using the ether **C–O** parameters. This therefore required the introduction of additional parameters, specific to acetal **C–O** bonds.

There is a way of testing, to some extent, the transferability of parameter values. Once the observables database has been assembled, some data may be retained and not included in the optimisation. The forcefield resulting from the optimisation may then be tested against the retained data to see if it is 'externally' consistent, and not only 'internally' consistent with the data it was fitted to.[7] In ideal circumstances,

agreement for the external data will be as good as for the internal data.

Choice of data for use in the optimisation must be made with the utmost care, since experimental errors in such data could be passed on to the forcefield. In practice, ideal data for optimisation can often be scarce, and so retaining some for use as external data may be an unaffordable luxury. In this case, it is often best to use the better data for optimisation, while retaining other, less accurate data as a check on transferability.

Finally, if optimised parameters are *not* found to be sufficiently transferable, it may be that the model compounds used are not representative of their class. It might then be necessary to re-appraise the observables database and add additional data from other model compounds.

2.3 The Observables Database

In assembling a database of experimental observables, there are two primary considerations. These are:

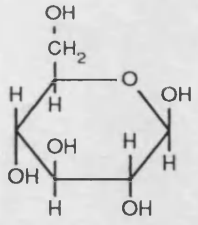
(i) Model Compounds to be included

(ii) Molecular Properties of interest

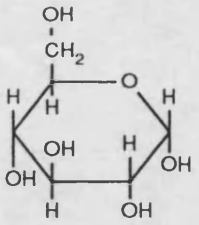
We shall look at each of these in turn.

2.3.1 Model Compounds

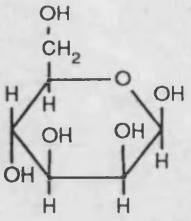
Model compounds were chosen that reflected the variety of structural features found in carbohydrates. A range of some of the more common carbohydrates are shown in Figure 2.2. Most large carbohydrate molecules are in the form of polysaccharides: pyranose and furanose residues linked together in chains.



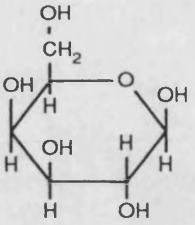
β -D-Glucose



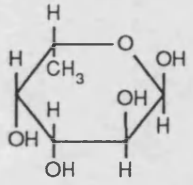
α -D-Glucose



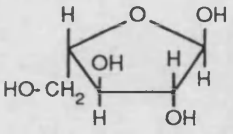
β -D-Mannose



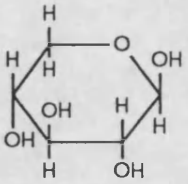
β -D-Galactose



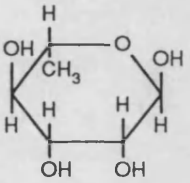
L-Fucose



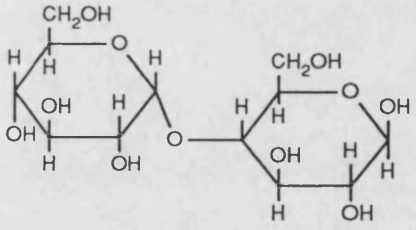
L-Arabinose



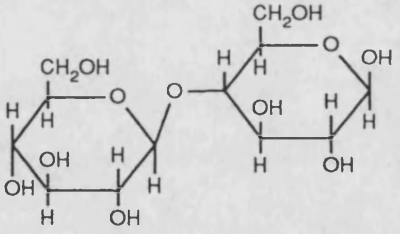
D-Xylose



L-Rhamnose



Maltose



Cellobiose

Figure 2.2 Some Common Carbohydrate Structures

Four classes of compounds were chosen as model compounds: hydrocarbons, ethers, alcohols and acetals. Because of the preponderance of ring systems within carbohydrates, model compounds containing five and six-membered rings were deemed to be particularly important model compounds. Most of the model compounds used in this study are shown in Table 2.1. The structures of non-trivial molecules, if not indicated in the text where they are referred to, are given in Appendix III at the end of this thesis.

2.3.2 Molecular Properties

In Section 2.1, four areas were identified as being of interest: molecular structure, conformational energies, molecular motion and intermolecular interactions. In order to be able to simulate these areas, data representative of each should be included in the observables database. The types of data selected are shown below.

Molecular Structure	Gas Phase Molecular Geometries (Electron diffraction, Microwave)
Conformational Energies	Conformational Energy Differences (NMR, various other techniques)
Molecular Motion	Vibrational Frequencies (IR & Raman), Rotational Barriers (IR & Raman, Microwave)
Intermolecular Interactions	Crystal Structures (X-ray Crystallography), Sublimation Energies (thermodynamic measurements)

The necessity of using such a wide range of data becomes apparent by considering the *Born-Oppenheimer surface*.[4,6] The conformational energy can be envisaged as a surface in multidimensional space, with each dimension representing one of the conformational degrees of freedom of the molecule. The potential energy function of the

Table 2.1 Model Compounds used in the Derivation of the Forcefield

Class	Data Type	Model Compounds
Hydrocarbons	Internal Geometry	Ethane, Propane, *n*-Butane, *i*-Butane, Cyclohexane, Neopentane, Cyclopentane, tri-(*t*-Butyl)-methane.
	Vibrational Frequencies	Ethane, Propane, *n*-Butane, *i*-Butane, Cyclohexane.
	Rotational Barriers	Ethane, *n*-Butane, 2-Methylbutane, 2,2-Dimethylbutane.
	Conformational Energies[a]	Cyclohexane, Methylcyclohexane, Decalin, 1,4-Dimethylcyclohexane.
	Crystal Structure	Cyclohexane, *n*-Octane.
Ethers	Internal Geometry	Dimethylether, 1,4-Dioxane, Ethylmethylether, Tetrahydrofuran, Tetrahydropyran.
	Vibrational Frequencies	Dimethylether, Diethylether, 1,4-Dioxane, Ethylmethylether, 1,2-Dimethoxyethane.
	Rotational Barriers	Dimethylether, Ethylmethylether, Diethylether, *i*-Propylmethylether, *t*-Butylmethylether, 1,2-Dimethoxyethane.
	Conformational Energies[a]	Diethylether, Methoxycyclohexane, *cis* and *trans*-2-Methoxy-*cis*-decalin, 2,2-Dimethylmethoxycyclohexane.
	Crystal Structure	Diethylether (phases I and II), 1,4-Dioxane.
Alcohols	Internal Geometry	Methanol, Ethanol, *i*-Propanol.
	Vibrational Frequencies	Methanol, Ethanol.
	Rotational Barriers	Methanol, Ethanol, *i*-Propanol, *t*-Butanol.
	Crystal Structure	Ethanol.
Acetals	Internal Geometry	Dimethoxymethane, 1,3-Dioxane, 1,3,5-Trioxane, Paraldehyde, 2,2-Dimethoxypropane.
	Vibrational Frequencies	1,3,5-Trioxane.
	Conformational Energies[a]	Dimethoxymethane, 2-Methoxytetrahydropyran, 2-Methoxy-1,3-dioxane, 2,4,6-Trimethyl-1,3-dioxane, 2-Methoxy-4-methyltetrahydropyran, 2-Methoxy-6-methyltetrahydropyran, *cis*- and *trans*-1,8-dioxadecalin.
	Crystal Structure	Trioxane.

a Also includes *Configurational* energy differences.

forcefield is itself a representation of the Born-Oppenheimer surface, and by fitting the forcefield to experimental data, we are seeking to make this representation as accurate as possible.

Because the Born-Oppenheimer surface cannot be directly determined, we must rely on experimental data to provide information about it. Thus, the molecular geometry of a conformer gives the location of an energy minimum, while a conformational energy difference defines the relative 'heights' of two such minima. Vibrational frequencies depend on the second derivatives of the surface around a minimum, and as a result give information on the *curvature* of the energy surface at that point. Finally, energy barrier heights separating two conformations may be obtained from rotational barrier measurements.

2.4 Experimental Data

Having decided on the molecular properties to be used in fitting the forcefield, it is necessary to consider the types of data available, and the techniques used in measuring them.

Some general principles in the selection of data can be made. Because calculations are generally performed on isolated molecules, experimental data should be selected with this in mind. Thus, data obtained from gas phase experiments will generally be most suitable, but where this is not available (as is often the case), data from measurements on dilute solutions in non-polar solvents are also acceptable. Data from polar solutions, or from measurements on the solid phase, are sometimes used where no other data can be obtained, but it should be emphasised that these may contain the effects of strong intermolecular interactions that will not be accounted for in the calculation.

2.4.1 Experimental Molecular Structure

Currently, there are four experimental techniques that are widely used for determining molecular structure. These are x-ray and neutron diffraction, which are carried out on crystals, and electron diffraction and microwave spectroscopy, which are carried out on gases. For the reasons given above, the structures derived from the latter two gas phase methods are preferred for parameterisation of the forcefield.

The definition of molecular geometry is, unfortunately, not a simple one. Both electron diffraction and microwave spectroscopy give quantities that are nominally referred to as bond lengths and angles - but these techniques actually measure different physical quantities. Not surprisingly therefore, bond lengths and angles are often slightly different depending on the method of measurement used. From electron diffraction, the intermolecular distances obtained are generally labelled r_a, r_g, r_α or r_α^0. Microwave spectroscopy, on the other hand, gives quantities labelled r_0 or r_s. A further quantity, r_e, is also occasionally derived from either of these methods.

Unfortunately, there are no simple general corrections that allow conversion between these structure types to be made. This raises a question as to which of these structure types is most appropriate for comparison with calculations.

The various structure types can be best understood by considering the effects of thermal vibration on the molecular geometry. The vibrational motion of any two bonded atoms is described by a Morse curve. This curve is close to a parabola at the minimum but at short distances the energy rises more steeply, at and long distances more slowly (as shown below). This means that as the temperature is increased, the

vibrational energy of the bond is raised, and it tends to get longer.[8]

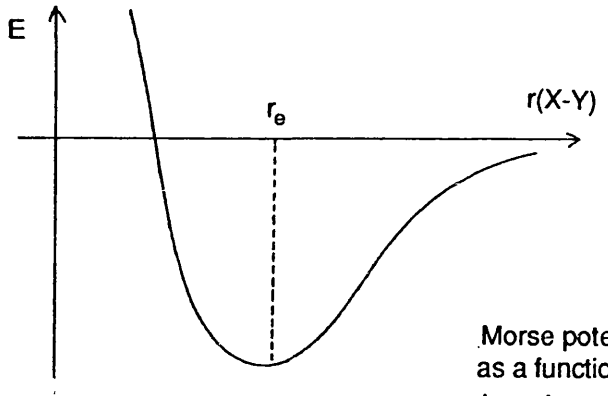


Figure 2.3

Morse potential representing the Bond Energy as a function of the bond length (r) between two atoms (X and Y).

The r_e structure is perhaps the easiest to understand: this is the internuclear distance corresponding to the hypothetical circumstance where each nucleus is at the bottom of its potential well. (It is a hypothetical situation because the vibrational energy of a bond is quantised, and cannot fall below its ground state.) The r_e structure would be the most desirable type of data for optimising forcefield parameters, but so few have been determined that their use as the only source of data for this purpose is out of the question.[8]

Geometries from Electron Diffraction

Electron diffraction makes use of the fact that electrons are scattered when passing between two nuclei, and that the degree of scattering is dependent on the internuclear distance. Electron diffraction can therefore be used to give a direct measurement of the internuclear distances within the molecule, in the form of a *radial distribution function.* In order to deduce the molecular structure, it is then a matter of finding the geometry that best fits this function. The geometry obtained in this manner is called the r_a structure. This may be converted into the r_g (which is in practice almost identical to r_a) by averaging over all of the molecular vibrations. The r_g structure can

be regarded as the thermal average of the internuclear distance.[7]

A further structure, r_α, is sometimes derived from the r_g geometry, and can be understood as follows. The two atoms in a bond each have an equilibrium position, and can vibrate in two ways: along the line joining these positions, and also perpendicular to it. Both of these vibrations contribute to the r_g value for the bond length. By applying a correction term, the component to the bond length arising from the perpendicular vibration can be removed, resulting in the r_α bond length.[9] Using a further correction, the value of r_α extrapolated to 0 K (denoted r_α^0) may be obtained. This can be regarded as the average geometry of the molecule in its vibrational ground state.

Because of the temperature effects on bond length, described above, bond lengths determined by electron diffraction are usually in the following order of magnitude:

$$r_e < r_\alpha^0 < r_\alpha < r_a \approx r_g$$

As an example, for **C–C** bonds, r_α^0 bond lengths are typically about 0.002 Å shorter than r_g values.[7] Because too few r_e structures are available, the preferred values for forcefield optimisation are r_α^0 (as r_α, r_g and r_a are all to an extent temperature dependent). In practice, however, r_α^0 structures are not always obtainable and so the r_α, r_g and r_a values sometimes have to be used.

A further point to note is that in some electron diffraction studies, bond lengths that are similar (but inequivalent) can present problems. This is because they appear so close together on the radial distribution function that they become difficult to resolve accurately. A test for this can be made by considering the *correlation coefficient* found between the two bond lengths in the fit of the distribution function. The correlation coefficient will be close to one (or 100%) if the resolution is poor.

It is because of resolution problems that equivalence of certain bond lengths is often assumed when electron diffraction structures are determined. A common assumption, for instance, is that all **C–H** bonds in a molecule are the same. It is

important that all such assumptions that have been made in solving the structure are born in mind when comparing with a calculated structure.

Geometries from Microwave Spectroscopy

Microwave spectroscopy operates by exciting the rotational energy levels of the molecule, and can be used to determine the three rotational constants, A_0, B_0 and C_0, of the vibrational ground state.[9] A structure, denoted r_0, may then determined that gives the best fit to the three rotational constants. However, because microwave spectroscopy measures only three quantities, in principle only three molecular structural parameters may be determined. Generally the method is therefore restricted to small molecules, and assumptions have to be made to reduce the number of independent structural parameters to three. These assumptions can be as to the equivalence of certain parameters, or by assigning fixed values for them. The number of independent structural parameters will also be reduced by equivalence due to symmetry, and so molecules chosen for microwave studies tend to be symmetrical.

These limitations can be overcome to some extent by making various substitutions of the component nuclei with different isotopes. The different atomic masses will result in a new set of rotational constants for each of the isotopically substituted derivatives made. This extra data allows more structural parameters to be evaluated by using the *Kraitchman Equations*.[10] The structure derived from isotopic substitution is called the r_s structure.

Structures determined by microwave spectroscopy, because of the assumptions often made in determining them, are only used for optimisation when electron diffraction data is not available.

2.4.2 Experimental Vibrational Frequencies

There are two experimental methods for measuring vibrational frequencies; Infrared (IR) and Raman spectroscopy. Each vibrational frequency value can be assigned to a particular *normal mode* of the molecule, and because the selection rules for infrared and Raman are different, the two techniques are often used to complement each other to make these assignments more reliable.[11]

Vibrational frequencies are an important source of experimental data because they give the most direct indication of the values of the force constant parameters in the potential energy function. This relationship between force constants and frequencies is described in more detail in the next chapter, where the calculation of vibrational frequencies is discussed.

One of the problems of vibrational spectra is that they give almost *too much* information. Because a molecule with N atoms gives rise to $3N-6$ vibrational modes, even medium-sized molecules can often have so many frequencies that it can be difficult to determine the correspondence between experimental and calculated frequency values. For use with forcefield parameterisation, small molecules are therefore better, as they have fewer vibrational modes. The problem may be further reduced by choosing model compounds of high symmetry. Symmetric molecules give rise to vibrations each having a particular *symmetry species*; this indicates which of the symmetry elements within the molecule are preserved by the vibration (see Chapter 3). Because each experimental and calculated frequency now has an associated symmetry species, finding the correspondence between them is simplified. It is for exactly the same reasons that small symmetrical molecules have traditionally been the subject of conventional vibrational analysis.[11,12] A fortunate result of this, from our point of view, is that vibrational data suitable for parameterisation is both reasonably abundant and well-assigned.

2.4.3 Experimental Rotational Barriers

Methods for measuring barriers to internal rotation are much more diverse than those for either molecular geometry or vibrational frequencies. Techniques that have been used include calorimetric measurements, variable temperature dipole moment studies, estimates from vibrational spectroscopy, microwave spectroscopy, and even sound absorption.[13] The two most common, and most accurate methods used are vibrational and microwave spectroscopy.

Vibrational spectroscopy can be used to give an estimate of the barrier height from the frequency values of the torsional vibrational modes. The torsional modes are generally found in the far-infrared region of the spectrum, and can sometimes be difficult to observe. Once a torsional frequency has been obtained, however, is is relatively simple to estimate a barrier height by assuming a simple mathematical form for the barrier.[14] Gas phase IR studies normally give barrier heights accurate to within 10-15% when compared with more accurate microwave methods. Errors are typically slightly larger for Raman studies (10-20%) since most are made on liquids or solutions and are affected by intermolecular interactions.[14]

Microwave spectroscopy generally gives the most accurate values for rotational barriers, with errors of about 5%.[13] It has the additional advantage of relating to molecules in the gas phase. The most common microwave method used is the *splitting method*, but this is generally restricted in its application to the rotational barriers of methyl groups (or other 'symmetric-top rotors' such as *t*-Butyl groups) for reasons of symmetry.[14]

2.4.4 Experimental Conformational Energies

Methods of determining conformational energy differences generally depend on establishing the relative populations of molecules in each of the conformations. For an equilibrium between conformations, *A* and *B*, the free energy difference between

them is given by the relationship:

$$\Delta G_{A \rightarrow B} = -RT \ln \left[\frac{n_A}{n_B} \right]$$

where n_A and n_B are the mole fractions of *A* and *B* respectively. By studying the equilibrium over a temperature range, the entropic component to ΔG can be eliminated, giving the enthalpy difference ΔH between the conformations.

The conformational energy difference can be in principle determined by any experimental method that can distinguish between the two conformations and measure their relative abundance. By far the most common method used, however, is NMR spectroscopy. This is because the NMR spectrum is generally readily interpretable in terms of the two conformations, and an accurate ratio of the populations may be obtained by integration. NMR spectra are normally run on dilute solutions in deuterated solvents; results from non-polar solvents (e.g. CCl_4, CS_2) being the most suitable for our purposes.

There are some restrictions on NMR methods, however. In order to determine the relative conformer populations, it is necessary that the conformational energy difference is around 2 kcal/mol or less, so that both conformers are present in observable quantities. It is also important that the barrier between conformers be higher than about 5 kcal/mol so that exchange between them is slow on the NMR time-scale (otherwise only a time-averaged spectrum of the molecule will be obtained).

2.5 References to Chapter 2

1. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).

2. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).

3. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).

4. O. Ermer, *Structure and Bonding*, **27**, 161, Berlin (1976).

5. S. Melberg and K. Rasmussen, *J. Mol. Struc.*, **57**, 215 (1979).

6. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).

7. K. Rasmussen, *Potential Energy Functions in Conformational Analysis,* in *Lecture Notes in Chemistry, Vol 37*, Springer-Verlag, Berlin & Heidelberg (1985).

8. U. Burkert and N.L. Allinger, in *Molecular Mechanics, ACS Monograph 177*, American Chemical Society, Washington, D.C. (1982).

9. K. Kuchitsu and K. Oyanagi, *Faraday Discuss. Chem. Soc.*, 21 (1976).

10. J. Kraitchman, *Amer. J. Phys.*, **21**, 17 (1953).

11. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations,* McGraw Hill, New York (1955).

12. P. Gans, in *Vibrating Molecules*, Chapman & Hall, London (1971).

13. E. Bright Wilson, Jr., *Adv. Chem. Phys.*, **II**, 365 (1959).

14. J. P. Lowe, *Prog. Phys. Org. Chem.*, **6**, 1-80 (1968).

Chapter 3

Calculation of Molecular Properties

3.1 The Potential Energy Function

One of the basic principles of conformational analysis is that for a given molecule, some geometrical arrangements (or conformations) of its atoms will be lower in energy than others. The most convenient way to express any conformation is in terms of *internal coordinates* - that is, by specifying the various bond lengths, valence angles and torsion angles of the molecule.[1] In order to gain some insight into the relationship between molecular geometry and conformational energy, a useful theoretical tool would be a mathematical function that can be used to calculate the conformational energy from the internal coordinates. The difficulty, of course, lies in determining such a function.

From our knowledge of molecular structure, we know that bond lengths, valence angles and torsion angles often have similar values from one molecule to another; and only in strained molecules (of high energy) do they deviate by much. This leads to the concept that these bond lengths, valence angles and torsions have certain *preferred* values, and that the conformational energy of the molecule increases depending on how much they are distorted from these values. In an attempt to quantify this concept, a fairly natural progression is to attempt to find a potential energy function that relates the conformational energy to the three types of variables: the bond lengths (denoted b), the valence angles (θ) and the torsion angles (ϕ) present in the molecule.

A fourth type of variable is also required in the potential energy function, and that is interatomic distance (denoted r_{ij} - the distance between atoms i and j).

Interatomic distances are included because of two other effects that also contribute to the overall conformational energy; the van der Waals and electrostatic interactions. These interactions are dependent on the *distance* between atoms not directly linked by bonds, hence the need for the r_{ij} variables.

All molecular mechanics forcefields use potential energy equations that are functions of the four variable types b, θ, ϕ, and r_{ij}; but the precise form of the function can vary widely. The functional form used in the present work is an extension of that used in the VFF, a forcefield originally developed, and currently used, for peptides and proteins.[2] Some additions have been made in order to account for aspects of the conformational behaviour of carbohydrates that the VFF could not reproduce. These additions will be indicated in the following discussion of the energy function.

The full form of the potential energy function is given in Equation 3.1

$$E = \sum \left\{ K_b \left[1 - \exp(-\alpha\{b - b_0\}) \right]^2 - K_b \right\} \qquad \text{Bond Strain Energy}$$

$$+ \tfrac{1}{2} \sum K_\theta (\theta - \theta_0)^2 \qquad \text{Angle Strain Energy}$$

$$+ \sum V_1 (1 + \cos\phi) + \sum V_2 (1 - \cos 2\phi) + \sum V_3 (1 + \cos 3\phi) \qquad \text{Torsional Energy}$$

$$+ \sum\sum K_{bb'} (b - b_0)(b' - b_0') \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum\sum K_{\theta\theta'} (\theta - \theta_0)(\theta' - \theta_0') \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum\sum K_{b\theta} (b - b_0)(\theta - \theta_0) \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum\sum\sum K_{\phi\theta\theta'} \cos\phi\, (\theta - \theta_0)(\theta' - \theta_0') \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum\sum K_{b\phi} (b - b_0)(1 - \cos 2\phi) \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum\sum -K_{b'\phi} (b' - b'_0)(1 - \cos 2\phi) \qquad \left.\right\} \text{Cross Terms}$$

$$+ \sum \varepsilon_{ij} \left[\left(\frac{r^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^*_{ij}}{r_{ij}} \right)^{6} \right] \qquad \text{Van der Waals Energy}$$

$$+ \sum \frac{q_i q_j}{r_{ij}} \qquad \text{Electrostatic Energy}$$

Equation 3.1

This equation relates the conformational energy (E) to the four variable types discussed above. All the other quantities present in the function (K_b, b_0, θ_0, K_θ, r^*_{ij}, q_i etc.) are constants called the forcefield *parameters*. Parameter values are selected by fitting properties calculated by the forcefield to experimental data, and are of critical importance to the performance of the forcefield. (The final values of the parameters for carbohydrates are given in Appendix I).

In Equation 3.1, a description of each of the terms in the function is indicated. The interpretation of each of these terms will now be considered in detail.

3.1.1 Bond Strain Energy

$$E_{Bond} = \sum \left\{ K_b \left[1 - \exp(-\alpha\{b - b_0\}) \right]^2 - K_b \right\}$$

The bond stretch energy is represented by an exponential ('Morse') function. The parameter b_0 represents the preferred bond length i.e. the bond length at which no bond strain occurs. The parameters K_b and α together affect the 'stiffness' of the bond - the force required to stretch or compress the bond by a given amount. For this reason, the quantity K_b is often referred to as the bond-stretch force constant.

In some simpler forcefields,[3-5] the bond strain is represented by a simple harmonic function

$$E_{Bond} = \sum \tfrac{1}{2} K_b' (b - b_0)^2$$

This representation should be sufficient as long as the distortion of the bond is small, but at larger distortions one would expect deviations from the harmonic potential. In reality, at very long distortions, the bond will start to dissociate, and the energy will

not rise above a certain value (the dissociation energy). In the harmonic approximation however, this will not occur and the energy will go on increasing with bond length. The morse function used in our potential should give a better representation, since it should reproduce the bond-weakening at longer bond lengths. The difference between the morse function and the harmonic function is shown in Figure 3.1.

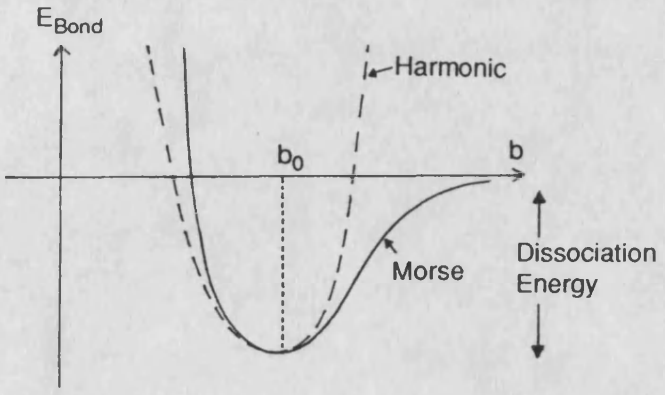


Figure 3.1

3.1.2 Angle Strain Energy

$$E_{Angle} = \tfrac{1}{2} \sum K_\theta(\theta - \theta_0)^2$$

The harmonic approximation appears to be a reasonable one in the case of angle bending, and so it has been retained in our forcefield. The parameter θ_0 is analogous to b_0 in that it represents the 'strain-free' bond angle. The parameter K_θ is the force constant for angle bending, and is related to the stiffness of the angle.

3.1.3 Torsional Energy

$$E_{Torsion} = \sum V_1(1 + \cos\phi) + \sum V_2(1 - \cos 2\phi) + \sum V_3(1 + \cos 3\phi)$$

The torsional energy term is represented by a Fourier expansion containing three terms - a onefold term (V_1) a twofold term (V_2) and a threefold term (V_3). As a first approximation, the threefold term should be sufficient for most single bonds, since it gives maxima at the eclipsed positions (0°, 120°, -120°) and minima at the staggered positions (60°, 180°, -60°). However, a non-zero value of the onefold parameter (V_1) was found necessary on occasion to reproduce experimental results. In general, a onefold term is used to represent a dipole-dipole interaction.[6] The twofold term, although not used in any of the calculations in the present work (V_2 = 0.0 for all single bond torsions parameterised so far) it is included here for completeness and because it is necessary for the treatment of double and conjugated bonds. (A further energy term, the 'out-of-plane' term is also required for conjugated systems.[2] This represents the energy required to distort the conjugated system from planarity. Although it has been maintained in the present forcefield, it is not discussed here because no conjugated systems have been studied in this work.)

The three-part functional form for the torsion angle is a departure from the VFF forcefield, which allowed only one such term per torsion - *either* onefold, twofold *or* threefold.[2]

3.1.4 Cross Terms

In addition to the previous terms described (which are also referred to as *diagonal* terms) the energy function also includes *off-diagonal* or *cross terms*, which represent coupling between deformations of two or more internals. For example, the bond-bond coupling term defines the additional energy required to stretch a bond (b) when an adjacent bond (b') is already stretched. We include the following coupling terms: bond-bond (with a common atom), angle-bond (the bond is part of the angle), angle-angle (with a common bond), and torsion-angle-angle (with the two angles involved in the torsion). The parameters in the cross terms denoted K (with the relevant

subscripts) are the cross term force constants. A further two cross terms are included that were not in the VFF forcefield; the two bond-torsion cross terms (relating b and ϕ, and b' and ϕ). These were included to reproduce the anomeric effect, and are only necessary for the anomeric **C–O** torsions in acetals and hemi-acetals. A full description of these cross terms is given in Chapter 7.

3.1.5 Van der Waals Energy

$$E_{VDW} = \sum \varepsilon_{ij} \left[\left(\frac{r^*_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r^*_{ij}}{r_{ij}} \right)^{6} \right]$$

The van der Waals energy is represented in the forcefield by a sum of pairwise Lennard-Jones interactions. The interactions are calculated between atoms separated by three bonds or more. The Lennard-Jones function was originally proposed to explain the interatomic forces occurring in noble gas atoms,[7] and comprises of two separate terms; an attractive term, dependent on r_{ij}^{6}; and a repulsive term dependent on r_{ij}^{n}. (*n* typically has values from 9 to 12. For mathematical convenience, the value $n=12$ is most often used.) The r_{ij}^{6} term accounts for attractive forces caused by polarisation effects; while the r_{ij}^{12} term relates to repulsions caused by nuclear and electronic repulsion. At short distances, the repulsive term dominates, while at longer distances the attractive term dominates. The van der Waals potential resulting from the Lennard-Jones function is shown in Figure 3.2

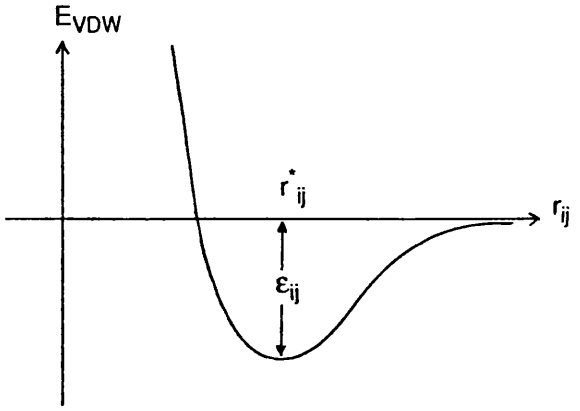


Figure 3.2 The Van der Waals Energy Function

The parameters r^*_{ij} and ε_{ij} represent the equilibrium distance and minimum energy respectively of the Lennard-Jones function. As in previous studies[8] the values of r^*_{ij} and ε_{ij} are estimated from the parameters for the corresponding homoatomic interactions; thus

$$r^*_{ij} = \tfrac{1}{2}(r^*_{ii} + r^*_{jj}) \; ; \quad \text{and} \quad \varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}.$$

3.1.6 Electrostatic Energy

The electrostatic energy of the molecule is approximated as a sum of pairwise Coulombic interactions between point charges centred on the atoms.

$$E_{Elec} = \sum \frac{q_i q_j}{r_{ij}}$$

The parameters q_i and q_j are the partial charges on the atoms i and j. For the calculations made in this work, the values of q_i and q_j are fixed for all atom types except carbon atoms. The charge of a given carbon atom depends on its local environment and is selected to maintain overall neutrality. A fuller discussion of partial charges is made in Chapter 5, which looks in detail at the performance of the non-bond part of the forcefield. (The charge parameters are also given in Appendix I.)

Many other forcefields use a Coulombic representation of electrostatic interactions, although some divide the electrostatic energy by a dielectric constant, D.[3]

$$E_{Elec} = \sum \frac{q_i q_j}{D\, r_{ij}}$$

In some cases, D is a function of r_{ij}, and is referred to as a *distance dependent* dielectric.[4,5]

We do not use a dielectric constant, because it is not clear that the concept of a bulk dielectric constant applies in a molecule.[9] It should also be noted that, other than for calculating the dipole moment, a simple numerical dielectric constant is equivalent to scaling the charge parameters q_i by $D^{-1/2}$, and that the same effect could be achieved by using smaller q_i values in the first place.

3.2 Energy Minimisation

Having defined the potential energy function, a wide range of properties of the system can in principle be calculated for comparison with experimental values. However, because most experimental data relate to molecules in conformational minima, a method of locating corresponding calculated minima is required. This can be achieved by minimisation of the potential energy function (Equation 3.1) with respect to the atomic coordinates of the molecule. Computationally, this is most readily performed in Cartesian coordinates, even though the potential energy is described in terms of internal coordinates. This therefore necessitates a coordinate transformation, which has been described elsewhere.[10]

From Equation 3.1, the first and second derivatives of the energy with respect to the internal coordinates can be derived. By a coordinate transformation, the derivatives with respect to the Cartesian coordinates may therefore be obtained. These can then be used to minimise the energy of the molecule with respect to all the atomic positions. For a molecule of *N* atoms, the geometry can be expressed by a set of *3N* Cartesian coordinates. The energy of the molecule will be at a minimum when the first derivatives of the energy is zero for all of the *3N* coordinates, thus:

$$\frac{\partial E}{\partial x_i} = 0 \qquad i=1,3N$$

Equation 3.2

Minimisation methods are concerned with locating the set of atomic coordinates for which this criteria is met, and invariably work by a process of iteration. The various minimisation methods differ in the way in which they achieve zero first derivatives. In practice, minimisation is performed until all the first derivatives fall below a specified (small) value. The smaller this value, the more fully minimised are the coordinates.

The three minimisation methods that have been used in the calculations reported in this work are the steepest descent, the Newton-Raphson, and a *quasi*-Newton method. A description of each, together with its applications, is given below.

3.2.1 Steepest Descent Method

The steepest descent method is one of the simplest minimisation methods used. It makes the assumption that the first derivatives ($\partial E/\partial x_i$) will be proportional to the displacement of the coordinates x_i from their values at the minimum. In effect, this assumption is an approximation of E as a quadratic function of the displacement of x_i. Given a starting geometry, denoted as vector **x** (containing the *3N* elements x_i) the first derivatives ($\partial E/\partial x_i$) can be calculated. A new estimate of the minimum geometry (**x′**) can then be obtained as follows:

$$\mathbf{x}' = \mathbf{x}+\delta\mathbf{x}$$

where the elements of vector $\delta\mathbf{x}$ are given by

$$\delta x_i = -L\left(\frac{\partial E}{\partial x_i}\right)_{\mathbf{x}}$$

The calculation can then be repeated for the new coordinates (**x′**) and the process iterated until Equation 3.2 is satisfied.

The quantity *L* in the above equation is a scaling constant, and in the algorithm we use it is dependent on the average magnitude of all *3N* derivatives ($\partial E/\partial x_i$).[11] As a result, when **x** is close to the minimum geometry, the derivatives will be small, and so will be the step size.

The steepest descent method performs very fast iterations because only first derivatives need to be computed, as opposed to other methods which also require second derivatives. It is particularly useful at geometries far from the minimum, where convergence is very rapid. As the geometry approaches the minimum, however, convergence becomes much slower, and other methods are preferred.

The steepest descent method was used in our calculations for highly strained molecules, and where the initial geometry might not have been close to the minimum. Examples of these are those compounds containing five-membered rings, or a high degree of steric crowding, such as tri-(*t*-Butyl)-methane. Steepest descent was used for the first 10 to 20 iteration steps, in order to relieve the large initial strain, and the quasi-Newton method then used for further refinement of the structure.

3.2.2 Newton-Raphson Method

The Newton-Raphson method is based on a Taylor series expansion of the energy around the minimum geometry.[9] The coordinates of the minimum can be expressed as a vector, $\mathbf{x}^0$, whose elements are the *3N* atomic Cartesian coordinates (x_i^0). At an initial starting geometry, ($\mathbf{x}^0 + \mathbf{x}$), the Taylor series is therefore:

$$E(\mathbf{x}^0 + \mathbf{x}) = E(\mathbf{x}^0) + \sum_{i=1}^{3N}\left(\frac{\partial E}{\partial x_i}\right)_{\mathbf{x}^0} x_i + \tfrac{1}{2}\sum_{i=1}^{3N}\sum_{j=1}^{3N}\left(\frac{\partial^2 E}{\partial x_i \partial x_j}\right)_{\mathbf{x}^0} x_i x_j + \ldots\ldots$$

Equation 3.3

Since the initial coordinates ($\mathbf{x}^0 + \mathbf{x}$) are known; the minimum can be located by finding the elements of $\mathbf{x}$ (the quantities x_i and x_j) from Equation 3.3.

The Newton Raphson method truncates the above Taylor series after the second derivative term, assuming that the higher order terms will be negligible near the minimum. This truncation is in effect an approximation of E as a quadratic function of $\mathbf{x}$. In matrix notation, the truncated form of the Taylor series becomes

$$E(\mathbf{x}^0 + \mathbf{x}) = E(\mathbf{x}^0) + \mathbf{g}\cdot\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x}$$

Equation 3.4

where $\mathbf{g}$ is the vector of $3N$ first derivatives ($\partial E/\partial x_i$); and $\mathbf{H}$ is the $3N \times 3N$ matrix containing the second derivatives ($\partial^2 E/\partial x_i \partial x_j$). The quantity $E(\mathbf{x}^0)$ is the energy of the minimum itself, and is therefore a constant.

Differentiating Equation 3.4 with respect to the vector $\mathbf{x}$ yields

$$\frac{\partial E}{\partial \mathbf{x}} = \mathbf{g} + \mathbf{H}\mathbf{x} = \mathbf{0}$$

since the elements of $\partial E/\partial \mathbf{x}$ will be zero at the minimum. Rearranging, we obtain

$$\mathbf{x} = -\mathbf{H}^{+}\mathbf{g}$$

Therefore by subtraction of the vector $\mathbf{x}$ from the initial coordinates, we should obtain a value for $\mathbf{x}^0$. However, because of the truncation of the Taylor series, this will only be an approximation of $\mathbf{x}^0$ (since E is unlikely to be an *exact* quadratic function of $\mathbf{x}$) and the process must be iterated to find the true minimum geometry.

The Newton Raphson method is most useful when the initial geometry is not too far from the minimum, and can be used to obtain very low first derivatives. The major

disadvantage, however, is that it is very slow, because the large second derivative matrix (**H**) must be first calculated and then inverted, which can take a considerable amount of time, especially for large molecules. The Newton-Raphson minimiser can also be problematical if the initial geometry is too far from the minimum.[6,11]

It was used in our calculations to refine geometries that had been already minimised by the quasi-Newton method, in order to give the accurately minimised structures (and second derivative values) that are necessary for the calculation of vibrational frequencies.

3.2.3 The Quasi-Newton Method

The quasi-Newton method is an adaptation of the Newton-Raphson method that does not require the exact calculation of the second derivative matrix **H** (see Equation 3.4). Instead, each iteration it forms an approximation of the matrix **H** from three sources: the value of **H** from the previous iteration, the difference between the first derivatives of successive iterations, and the step length vector, **x**. The value of **H** in the initial iteration is taken to be the identity matrix **E**. (For a full description of the quasi-Newton method, see reference 12)

Because the quasi-Newton method does not calculate the second derivatives analytically, it is much faster than the Newton-Raphson method, and is a very good general-purpose minimiser. The quasi-Newton method was used for all the calculations described in this work, either on its own, or in conjunction with the two minimisation methods discussed above.

3.3 Calculated Molecular Geometry

Because experimental geometries generally relate to molecules in their minimum energy conformation, the calculated molecular geometries used for comparison are obtained directly from minimisation. For calculated geometries, minimisations were

deemed complete when the maximum first derivative value was less than 1×10^{-5} kcal/mol/Å. The quasi-Newton minimisation method is sufficiently accurate for derivatives of this size.

For flexible molecules there may be several different minima: *n*-butane, for example, has three separate minima; one *trans* and two *gauche*. The minimum obtained by the calculation depends on the starting geometry used. A given starting geometry will generally only reveal one minimum - that which it is closest to.

Obviously, it is important that the calculated minimum is the same one as the experiment relates to. In practice, for small molecules there is rarely any confusion since nearly all the molecules in the experimental sample will be in the lowest minimum. For example, for an energy difference of 2 kcal/mol between conformations, 97% of the molecules will be in the lower energy conformer at room temperature. (The distribution of molecules between conformations can be estimated from the Boltzmann equation, $N_1/N_0 = e^{-\Delta E/RT}$.)

For the relatively small model compounds studied in this work, determining the lowest calculated energy is not a problem because only a few minima are feasible for each molecule. However, in larger, more flexible molecules, many local minima can exist, and an exhaustive search for the lowest minimum can present difficulties.

3.4 Calculated Conformational Energy Differences

Conformational energy differences were calculated by minimising the geometries of the two conformations in question. (The maximum derivative criterion was the same as that given above for molecular geometries.) The conformational energy difference then is simply the difference between the energies of each minimised conformation calculated from the potential energy function (Equation 3.1).

3.5 Calculated Rotational Barriers

The barrier to rotation about a single bond can be regarded as a saddle-point between two minima on the conformational energy surface. The shape of this energy barrier may in principle be determined by fixing the relevant torsion angle at a series of values, while allowing all the other internal degrees of freedom to relax. Since the minimisation is carried out in Cartesian coordinates, rather than internal coordinates, fixing an internal coordinate in this way would be a complex task. A similar effect, however, can be achieved more simply by using a method called *torsion forcing*. This method 'drives' the torsion angle through a range of values, minimising the geometry fully at each stage. The torsion angle stays at its 'fixed' value by means of an additional energy term - the forcing function - included in the forcefield. The forcing function takes the form of a harmonic potential

$$E_{Force} = K_F (\phi - \phi_F)^2$$

The value of K_F can be any large value sufficient to keep ϕ within a degree or so of ϕ_F. In our calculations, we used a K_F value of 1000 kcal/mol/rad^2. The value of E_{Force} can be subtracted from the total energy of the molecule at that particular ϕ value. In general, however, as K_F is so large, the deviation of ϕ from ϕ_F is very small, and so E_{Force} becomes negligible.

After each minimisation, the value of ϕ_F is incremented by 5° before the next minimisation begins. When all the minimisation steps have been completed, the resulting set of energy values can be plotted against the corresponding ϕ values to give a rotational barrier plot (see Appendix II for examples). The *height* of the barrier (the difference between the lowest and highest points on the plot) can then be compared with experimentally determined values for rotational barriers.

3.6 Calculated Vibrational Frequencies

The calculation of vibrational frequencies of a molecule is fairly straightforward if the Newton-Raphson method has been used for the minimisation, since the necessary second derivative matrix, **H**, is already available (Equation 3.4).

For small atomic displacements **x** from the minimum geometry, the kinetic energy (E_K) and potential energy (E_P) of a molecule are given in matrix notation by

$$E_K = \tfrac{1}{2}\dot{\mathbf{x}}^T\mathbf{M}\dot{\mathbf{x}} = \tfrac{1}{2}\dot{\mathbf{q}}^T\dot{\mathbf{q}} \qquad \text{Equation 3.5}$$

and

$$E_P = \tfrac{1}{2}\mathbf{x}^T\mathbf{H}^0\mathbf{x} = \tfrac{1}{2}\mathbf{q}^T\mathbf{M}^{-1/2}\mathbf{H}^0\mathbf{M}^{-1/2}\mathbf{q} \qquad \text{Equation 3.6}$$

M in the above equations is a diagonal matrix with the atomic masses as the diagonal elements: $\mathrm{diag}(\mathbf{M}) = m_1, m_1, m_1, m_2, m_2, m_2, m_3, \cdots, m_N, m_N, m_N$. The vector **q** is the mass weighted atomic displacements defined as

$$\mathbf{q} = \mathbf{M}^{1/2}\mathbf{x}$$

$\mathbf{H}^0$ corresponds to the second derivative matrix **H**, evaluated at the minimum ($H_{ij}^0 = (\partial^2 E_P/\partial x_i \partial x_j)_0$).

By inserting Equations 3.5 and 3.6 into Newton's equations of motion (see reference 13) the following equation is obtained:

$$(\mathbf{M}^{-1/2}\mathbf{H}^0\mathbf{M}^{-1/2} - \mathbf{E}\lambda)\,\mathbf{l} = 0$$

Nontrivial solutions are obtained only if the secular equation

$$|\,\mathbf{M}^{-1/2}\mathbf{H}^0\mathbf{M}^{-1/2} - \mathbf{E}\lambda\,| = 0$$

is satisfied (where **E** is the identity matrix). Solving this eigenvalue problem (by diagonalisation of the mass-weighted second derivative matrix, $\mathbf{M}^{-1/2}\mathbf{H}^0\mathbf{M}^{-1/2}$) yields the vibrational frequencies (the square roots of the eigenvalues λ_i) and the normal modes of vibration (the corresponding eigenvectors $\mathbf{l}_i$).

For a non-linear molecule with *N* atoms, there will be *3N* eigenvalue and eigenvector pairs; however, six of these will have eigenvalues of zero, corresponding to the three rotational and translational motions of the molecule. The molecule will therefore have *3N-6* remaining eigenvalue and eigenvector pairs, relating to the internal vibrations of the molecule. The eigenvectors indicate the atomic displacements occurring in each of the vibrations of the molecule. Transformation of the eigenvectors from Cartesian coordinates into internal coordinates facilitates the assignments in terms of molecular deformations, and such assignments are indispensable for properly matching observed and calculated vibrational frequencies.

The parameters that have the most influence on the second derivative matrix (**H**) - and hence the calculated vibrational frequencies - are the various force constant parameters (K_b, K_θ, $K_{bb'}$, etc.). This can be appreciated by a simple analogy with the case of a one-dimensional harmonic oscillator of mass M. The frequency of vibration, ν, depends on the force constant, K, as follows:

$$\nu = \frac{1}{2\pi}\left(\frac{K}{M}\right)^{1/2}$$

In a similar, though more complex way, the constant parameters are largely responsible in determining molecular vibrational frequencies.[13, 14]

3.6.1 Determining the Symmetry of Calculated Vibrations

In addition to the vibrational assignments described in the previous section, using symmetric model compounds greatly simplifies the matching of calculated frequency values with their corresponding experimental values. It does, however, require that the symmetry of the calculated vibrations be determined.

Molecules which are symmetrical in their minimum energy conformations give rise to vibrations which also display some of the symmetry present at the minimum. The *symmetry species* of the vibration indicates which of the symmetry elements present at the minimum are preserved during the course of the vibration.

As an example, we shall consider Dimethylether: in its minimum energy geometry, it has the point group C_{2V}. This point group has three symmetry elements (apart from the identity, E); which are a C_2 axis and two σ_V planes (see Figure 3.3). From a set of character tables (see, for example, reference 13) this point group can be found to have four symmetry species: A_1, B_1, A_2 and B_2. Figure 3.4 shows an example of a vibrational mode of dimethylether belonging to each of these symmetry species. A_1-type vibrations are totally symmetric and all the symmetry elements of the molecule are preserved throughout the vibration. For the other symmetry species, apart from the identity element E, only one of the symmetry elements is preserved: B_1 vibrations preserve only the $\sigma(xz)$ plane; A_2 only the C_2 axis; and B_2 only the $\sigma(yz)$ plane. The relationship between symmetry elements and symmetry species is given in character tables for all the possible molecular point groups.

In the course of this work, a program was developed for the purpose of determining the symmetry species of molecular vibrations. The program operates by taking the original coordinates of the molecule at the minimum ($\mathbf{x}_0$) and displacing them along the normal mode in question (using the relevant eigenvector, $\mathbf{x}$) :

$$\mathbf{x}' = \mathbf{x}_0 + \mathbf{x}$$

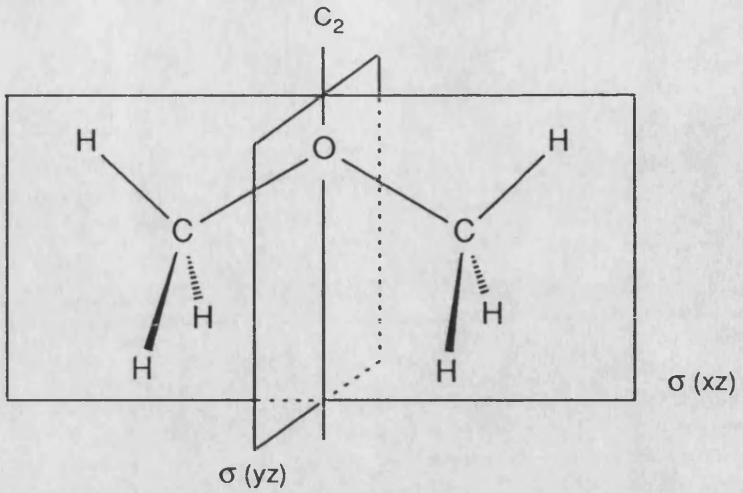
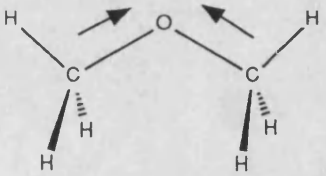


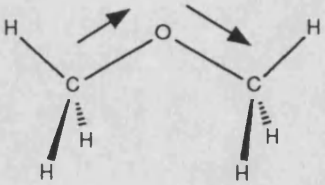
Figure 3.3 Symmetry Elements present in Dimethylether (point group C_{2V})

C-O Stretch



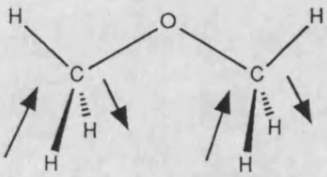
A_1 all symmetry preserved

C-O Stretch



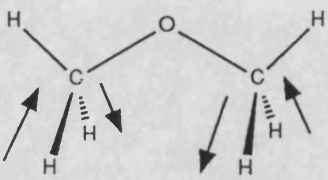
B_1 only σ(xz) preserved

C-H Stretch



B_2 only σ(yz) preserved

C-H Stretch



A_2 only C_2 preserved

Figure 3.4 Symmetry Species of Dimethylether Vibrations

The coordinates **x′** are then multiplied by the transformation matrix of one of the symmetry elements to generate some new coordinates **x″**. If the atoms of the transformed geometry, **x″** overlay those of the untransformed, **x′**; then the symmetry element has been preserved. However, if the symmetry element is not present in the displaced coordinates (**x′**), the geometries will not overlay each other exactly. The program tests each of the vibrational frequencies of the molecule for all of the symmetry elements present in the point group.

3.7 References to Chapter 3

1. E.L. Eliel, N.L. Allinger, S.J. Angyal, and G.A. Morrison, in *Conformational Analysis*, Interscience (1965).
2. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).
3. S. Melberg and K. Rasmussen, *J. Mol. Struc.*, **57**, 215 (1979).
4. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
5. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187 (1983).
6. U. Burkert and N.L. Allinger, in *Molecular Mechanics, ACS Monograph 177*, American Chemical Society, Washington, D.C. (1982).
7. P.W. Atkins, in *Physical Chemistry, 2nd Ed.*, Oxford University Press, Oxford (1982).
8. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).
9. S. R. Niketic and K. Rasmussen, in *The Consistent Force Field*, Springer, New York (1977).

10. A.T. Hagler, *Theoretical Simulation of Conformations, Energetics and Dynamics of Peptides,* in *The Peptides*, vol. 7, p. 213, Academic, New York (1985).

11. O. Ermer, *Structure and Bonding*, **27**, 161, Berlin (1976).

12. R. Fletcher, *Practical Methods of Optimization,* 1, Wiley, N.Y. (1980).

13. P. Gans, in *Vibrating Molecules*, Chapman & Hall, London (1971).

14. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations,* McGraw Hill, New York (1955).

Chapter 4

Determination of Forcefield Parameters

4.1 Optimisation Methods

The performance of a valence forcefield, that is, the reliability of its predictions, is critically dependent on the values of the parameters used. For this reason, a great deal of care must be taken in the selection of parameter values. The process of determining suitable forcefield parameters is often referred to as *parameterisation*.

The overall objective in deriving a forcefield is to find a *small* set of parameters which is sufficient to reproduce experimental results. As mentioned in Chapter 2, parameterisation may take one of two forms: optimisation by inspection,[1] or a rather more systematic *least-squares* optimisation method.[2,3] The latter method is certainly better when large numbers of parameters are being optimised to many observables simultaneously, as the relationship between parameters and observables can often be too complex for optimisation by inspection alone. This is not to say that least-squares optimisation is wholly without problems. Least-Squares methods work by minimising the deviation between calculated and experimental properties; which is a function of the parameter values. Just as minimising the energy of a molecule can result in different minima, depending on the starting geometry, so a least-squares fit can result in different optimised parameter sets, depending on the initial parameters used.

The preferred parameter set will be the one that gives the lowest overall deviation, but the only way of being sure that the optimal parameter set has been found would be to perform least-squares optimisations on every possible combination of initial parameter values. Obviously, this will not be possible, and in practice we must

use our own judgement to reduce the scale of the problem. This can take the form of selecting 'reasonable' initial parameter values, and by optimising parameters a few at a time, rather than attempting to optimise *all* parameters to *all* observables in one step.

The process of parameterisation is very labour intensive, and in the case of this work required a great many separate optimisation steps. The order in which parameters are optimised can also be of critical importance, and steps often have to be retraced several times over, so that different combinations of optimising the parameters can be tested. For this reason, a blow by blow account of the parameterisation will not be given here; instead, we shall first outline the least-squares method and then give a description of the general approach used.

Optimisation by least-squares methods, as indicated by the above discussion, is not perhaps as objective as it first appears. Another element of subjectivity occurs if more than one type of observable is being fitted at the same time. To optimise observables measured in different units, say bond lengths and valence angles, a judgement has to be made as to how much error in one property is equivalent to how much error in the other. A weighting scheme must therefore be applied, which allows deviations measured in different units to be properly compared. The weighting scheme chosen, also, to some extent, reflects the relative importance being attached to fitting each of the observable types.

4.2 The Least-Squares Method

Mathematically, the least-squares method can be formulated as follows.

Let $\mathbf{p}$ be a vector whose P components are the forcefield parameters p_j to be optimised; and $\Delta\mathbf{y}$ be a vector with the Y differences between the experimental data y_i^{obs} and their calculated counterparts y_i; thus $\Delta\mathbf{y} = \mathbf{y} - \mathbf{y}^{obs}$.

[Note that there are Y known values (y_i^{obs}) and P unknowns (p_j). P must be less than

or equal to Y, otherwise the problem will be underdetermined.]

The relationship between the y_i values and the p_j values can be expressed as a Taylor series, truncated after the linear term.

$$y_i(\mathbf{p}+\delta\mathbf{p}) = y_i(\mathbf{p}) + \sum_{j=1}^{P} \frac{\partial y_i}{\partial p_j}\delta p_j \qquad \text{Eqn. 4.1}$$

In matrix notation, this becomes

$$\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \mathbf{y}(\mathbf{p}) + \mathbf{Z}\delta\mathbf{p} \qquad \text{Eqn. 4.2}$$

where $\mathbf{Z}$ is the Jacobian matrix with elements $Z_{ij} = \partial y_i/\partial p_j$. Subtraction of $\mathbf{y}^{obs}$ (the vector containing the experimental data) from both sides of this equation gives

$$\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \Delta\mathbf{y}(\mathbf{p}) + \mathbf{Z}\delta\mathbf{p} \qquad \text{Eqn. 4.3}$$

(As $\Delta\mathbf{y}$ was defined above as $\mathbf{y}-\mathbf{y}^{obs}$.)

Since Equation 4.1 is a truncation of the Taylor series, and therefore only an approximation, $\delta\mathbf{p}$ represents only *improved* differences in $\mathbf{p}$, and not optimal ones. The process of determining $\delta\mathbf{p}$ must therefore be iterated until the condition

$$\sum_{i=1}^{Y}[\Delta y_i(\mathbf{p}+\delta\mathbf{p})]^2 = \text{minimum} \qquad \text{Eqn. 4.4}$$

is satisfied. (Equation 4.4 represents the sum of the squares of the deviations.) In vector notation, this becomes

$$\Delta\mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \text{minimum} \qquad \text{Eqn. 4.5}$$

If the observables are weighted (as discussed in the previous section) the function to be minimised is finally

$$\Delta\mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})\mathbf{W}\mathbf{W}\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \text{minimum} \qquad \text{Eqn. 4.6}$$

where **W** is a diagonal matrix with the weights of the observables. This function will be a minimum when the first derivative (with respect to $\delta\mathbf{p}$) is zero:

$$\frac{\partial}{\partial\delta\mathbf{p}_A}[\Delta\mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})\mathbf{W}\mathbf{W}\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p})]_{\mathbf{p}=\mathbf{p}_A} = \mathbf{0} \qquad \text{Eqn. 4.7}$$

($\mathbf{p}_A$ is the initial estimate of the parameters). Differentiating, using the product rule, we obtain

$$\left(\frac{\partial\Delta\mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})}{\partial\delta\mathbf{p}_A}\right)_{\mathbf{p}_A}\mathbf{W}\mathbf{W}\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) + \Delta\mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})\mathbf{W}\mathbf{W}\left(\frac{\partial\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p})}{\partial\delta\mathbf{p}_A}\right)_{\mathbf{p}_A} = \mathbf{0} \qquad \text{Eqn. 4.8}$$

Since the two terms in Equation 4.8 are equal, it may also be written

$$2\left(\frac{\partial \Delta \mathbf{y}^T(\mathbf{p}+\delta\mathbf{p})}{\partial \delta \mathbf{p}_A}\right)_{\mathbf{p}_A} \mathbf{WW}\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \mathbf{0} \qquad \text{Eqn. 4.9}$$

The partial derivative in this equation is equivalent to the transpose of the **Z** matrix evaluated at $\mathbf{p}=\mathbf{p}_A$ (since $Z_{ij} = \partial y_i/\partial p_j$). Thus

$$2\,\mathbf{Z}_A^T\mathbf{WW}\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p}) = \mathbf{0} \qquad \text{Eqn. 4.10}$$

Substituting $\Delta\mathbf{y}(\mathbf{p}+\delta\mathbf{p})$ from Equation 4.3; and also putting $\mathbf{W}^2 = \mathbf{WW}$ gives

$$2\,\mathbf{Z}_A^T\mathbf{W}^2(\Delta\mathbf{y}(\mathbf{p}) + \mathbf{Z}_A\delta\mathbf{p}) = \mathbf{0} \qquad \text{Eqn. 4.11}$$

Finally, rearranging to get $\delta\mathbf{p}$

$$\delta\mathbf{p} = -(\mathbf{Z}_A^T\mathbf{W}^2\mathbf{Z}_A)^{-1}\mathbf{Z}_A^T\mathbf{W}^2\Delta\mathbf{y}(\mathbf{p}) \qquad \text{Eqn. 4.12}$$

$\delta\mathbf{p}$ is the improved differences in **p**, and the new parameter set for the next iteration is then simply

$$\mathbf{p}_{new} = \mathbf{p} + \delta\mathbf{p} \qquad \text{Eqn. 4.13}$$

The time consuming part of the calculation is the computation of the **Z** matrix. The elements of this matrix are computed partly numerically and partly analytically, as described by Lifson and Warshel.[2] The **Z**-matrix elements are particularly useful in

themselves, because they give a quantitative measure of the influence of parameter p_j on the calculated observable y_i ($Z_{ij} = \partial y_i/\partial p_j$). A parameter can be removed from the optimisation if its influence on *all* y_i values is small enough, since it will be of little help in achieving the fit. The elements of the **Z**-matrix also give information about the relative importance of each parameter, and are especially useful in determining the order in which parameters should be optimised.

4.3 Parameters Included in the Optimisation

It should be initially stressed than the non-bond parameters (q_i, ε_{ii} and r_{ii}^*) were *not* optimised in this work. Instead, the values for these parameters were transferred from the VFF forcefield, where they had been originally derived from least-squares fitting to crystal properties.[4-6] A discussion of the non-bond parameters is made in the next chapter, where their applicability to carbohydrates and carbohydrate model compounds is demonstrated.

Parameters that were fitted in this study were the 'internal' parameters; that is, those relating to the energy terms of the forcefield equation (3.1) that depend on the internal coordinates, b, θ and ϕ. Initial values for the internal parameters were transferred from similar parameters in the VFF forcefield[7] in almost all cases. The cross term force constant parameters, however, were given initial values of zero, as were the values of V_1 and V_2.

All of the internal parameters were fitted, with the following exceptions:

(i) Parameters relating to the **C–H** bond (K_b, α, and b_0). These were left at their initial VFF-values, which had originally been taken from Lifson's hydrocarbon forcefield.[8]

(ii) The values of the α parameters were not fitted since they were found to be too highly correlated with the corresponding K_b values. (These parameters both affect bond 'stiffness'.) They were therefore left at their initial (VFF) values,[7]

and only the K_b values were optimised.

(iii) Where possible, cross terms were left at their initial zero values. They were only optimised when it was impossible to get a good fit to experiment without them.

4.4 Data Used for Optimisation

In principle, almost any experimental observable can be included in the least-squares optimisation process. In practice, however, we used only two types of data - molecular geometry and vibrational frequencies.

Conformational energy differences (and rotational barriers) were not used as least-squares observables for several reasons. Although calculation of these quantities is relatively simple (see Chapter 3) the construction of the **Z**-matrix presents special problems. This is because these observables relate to *differences* between two calculations, and therefore each **Z**-matrix element would have to 'belong' to two calculations. Rasmussen has also commented on the difficulties involved in calculating such a **Z**-element.[9] Other observables (for example a bond length, or a vibrational frequency) can be obtained from a single calculation and do not, therefore, give the same problem.

An equally important problem with conformational energies, however, is that there is generally only one such observable per molecule, dependent on a large number of parameter values. As stated in the discussion of the least-squares method, the number of observables has to equal or exceed the number of parameters for the parameter values to be determinable. Therefore, the inclusion of conformational energies in the least-squares optimisation will do little to assist in determining specific parameter values.

By contrast, molecular geometry and vibrational frequencies are ideally suited to parameter optimisation. There are generally many such observables per molecule, and the relationships between individual observables and specific parameter values is

reasonably straightforward. The procedure used in fitting the forcefield parameters was therefore to fit them in the first instance to molecular geometry and frequency values, and then to use conformational energies and rotational barriers as a check on the resulting parameter sets. Any large discrepancies could then be adjusted by the slight modification of relevant parameters, and occasionally the introduction of non-zero V_1 parameters (see Chapters 6 & 7).

Because only two types of data were used in the least-squares optimisation, the problem of weighting did not prove difficult to resolve. By inspection of the initial **Z**-matrix elements, it was found that the b_0 and θ_0 parameters were found to influence molecular geometry far more than the vibrational frequencies, while the opposite was true for the force constant parameters. The result was that, in most cases, b_0 and θ_0 could be fitted to structural data, and force constants to vibrational data, in separate optimisations. Thus, the need for weighting between vibrational frequencies and geometrical values did not arise.

In the fits to structural data, both bond lengths and bond angles were included in the same optimisations. In this case, the deviations in bond lengths were weighted by a factor of 250, so that deviations of 0.01 Å in a bond length were equivalent to deviations in valence angles of 2.5°.

In fitting a forcefield, it is obviously necessary to have in mind some criteria for judging whether of not the fit to experimental data is good enough. This is something that is very difficult to decide beforehand, since a feel for what can be achieved only comes with experience. After our initial fits to hydrocarbons, however, we considered a fit to be reasonable if the following criteria were met

Observable	*Maximum Deviation*
Bond length	0.015 Å
Valence angle	4°
Vibrational Frequency	50 cm^{-1}

Although these criteria are on occasion exceeded, they provided a frame of reference for the remainder of the forcefield development. It should be noted that *average* deviations are less than half the above values.

4.4.1 Fitting Vibrational Frequencies

Some aspects of fitting vibrational frequencies deserve special comment. For the optimisations to vibrational data, all frequencies were fitted except those above 2000 cm^{-1}. Frequencies above this are the **C–H** and **O–H** stretching frequencies; and these were not included because they are of little importance in determining conformational motions. The parameters for these bonds (K_b, b_0 and α) were therefore transferred from the VFF[7] and were not adjusted further.

All fitted frequencies were given a weighting of 1.0 in the least-squares optimisation process, apart from doubly degenerate frequencies which were each weighted 0.5 (since the frequency of degenerate modes, although applicable to more than one vibration, can still only be regarded as *one* observable).

In fitting vibrational frequencies, the molecules were first minimised using the initial parameter set and the normal modes calculated. The symmetries of the normal modes were then determined using the program described in Section 3.6.1. Knowing the symmetry species of the normal modes enabled the correspondence between the calculated normal modes and the experimental frequencies to be established.

As we have seen, the least-squares procedure works by an iterative method, changing the parameter values on each iteration. Each iteration therefore gives rise to a new set of normal modes slightly different to those of the previous iteration. Unfortunately, the least-squares program has no way of identifying a particular normal mode, other than by ordering the normal modes in order of frequency and using the place of the normal mode in this order to identify it. A problem arises however when two normal modes are close in frequency, as a small change in the potential parameters can reverse the order of their frequency values in the next iteration. This would have resulted in the fitting of the wrong normal modes to the observed frequencies and a mechanism for identifying normal modes that have 'crossed over' was therefore required.

The method developed relies on the premise that only small changes in parameter values will occur between successive iterations and so the normal modes generated will also change only slightly. It is therefore a matter of finding out which normal mode of the previous iteration is most similar in terms of molecular motion to each of the normal modes in the current iteration. This can be accomplished by considering the dot products between normal modes as follows. Normal modes of one iteration are all orthogonal to each other and the dot product between any two is therefore zero. Conversely the dot product of a normal mode with itself is unity. In the *k*th iteration therefore, for the normal mode $\mathbf{x}_i^k$:

$$\mathbf{x}_i^k \cdot \mathbf{x}_i^k = 1$$

$$\mathbf{x}_i^k \cdot \mathbf{x}_j^k = 0$$

It follows therefore that for a given normal mode; its dot product with the most similar normal mode from the previous iteration will be approximately one, while its dot product with all the other normal modes will be approximately zero.

$$\mathbf{x}_i^k \cdot \mathbf{x}_i^{k-1} \approx 1$$

$$\mathbf{x}_i^k \cdot \mathbf{x}_j^{k-1} \approx 0$$

Thus by determining the correspondence between the experimentally observed frequencies and the normal modes generated by the *initial* values of the potential parameters, the least-squares procedure will automatically ensure that the observed frequencies are fitted to the correct normal modes in successive iterations. Brady *et al.* have used a similar method to find a correspondence not between normal modes of successive iterations, but between normal modes calculated by different forcefields.[10]

4.5 Sequence of Optimisation

The optimisation process was made in several stages. Firstly, parameters involving only **C** and **H** were fitted to experimental data on hydrocarbons. These were then assumed to be transferable for use with ether molecules, and the ether model compounds were then fitted by optimisation of those parameters involving **O** only. The resulting parameters were then transferred to acetals, and the data for acetals were then fitted by optimisation of the anomeric parameters (those involving the **O–C–O** fragment) only. Alcohols were also fitted by the same process (i.e optimisation of the hydroxyl parameters only) although the ether **C–O** parameter b_0 was not found to be sufficiently transferable for use with alcohols (see Chapter 6) and a separate b_0 value for alcoholic **C–O** bonds was required.

The general strategy in fitting each class of compounds was first to optimise only the *diagonal* parameters (i.e. those pertaining to the diagonal terms, as opposed to the cross terms). Cross terms were only fitted were the diagonal terms alone could not reproduce experiment. They were found to be particularly necessary in the reproduction of vibrational frequencies, an observation also made by Ermer.[3]

4.6 Final Parameter Values

The final parameter values are given in Appendix I. All the calculations reported in the present work were performed using these parameter values.

4.7 References to Chapter 4

1. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).
2. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).
3. O. Ermer, *Structure and Bonding*, **27**, 161, Berlin (1976).
4. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979).
5. A. T. Hagler, S. Lifson, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5122 (1979).
6. A. T. Hagler, P. Dauber, and S. Lifson, *J. Am. Chem. Soc.*, **101**, 5131 (1979).
7. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).
8. A. T. Hagler, P. S. Stern, S. Lifson, and S. Ariel, *J. Am. Chem. Soc.*, **101**, 813 (1979).
9. K. Rasmussen, *Potential Energy Functions in Conformational Analysis,* in *Lecture Notes in Chemistry, Vol 37*, Springer-Verlag, Berlin & Heidelberg (1985).
10. S.N. Ha, A. Giamonna, M. Field, and J.W. Brady, *Carbohydr. Res.*, **180**, 207 (1988).

Chapter 5

Calculations on Crystals

5.1 Introduction

Calculations on crystals were made in order to evaluate the non-bond energy parameters. These were taken from previous calculations on alkanes,[1] amides and carboxylic acids,[2-5] in which they were optimised by a least-squares fitting procedure to reproduce crystal structures, dipole moments and sublimation energies.

The forcefield that developed from those earlier studies (the VFF[6]) is currently used by other workers in this laboratory in the study of peptide and protein conformation. Our reasons for not re-optimising the non-bond parameters for carbohydrates are twofold. Firstly they are found to give sufficiently good results for the selected model compounds as they stand (as outlined later in this chapter) and secondly to maintain consistency with the VFF. This consistency will enable us in future to incorporate the carbohydrate parameters and functional forms developed in this work into the VFF in order that protein-carbohydrate interactions may be modelled successfully.

Non-bond interactions are important because they are generally considered to be the determining factor in the conformation of large flexible molecules like carbohydrates. The non-bond parameters will therefore be critical to the performance of the forcefield when applied to these systems.

A good test of non-bond parameters is their ability to reproduce the structure and properties of crystals, as crystal packing is mainly determined by intermolecular non-bond interactions. Crystals also have two other particular advantages with regard to forcefield calculations: firstly, a vast range of experimental data is available,

as many crystal structures have been determined; and secondly, these structures are known to very high degree of accuracy. This accuracy means that the intermolecular interactions can be determined in detail, as the atomic positions within the crystal are known with great precision.

5.2 The Crystal Forcefield

The method used for modelling crystal structure in this work was the same as that reported by the Lifson group in their studies on carboxylic acids and amides.[3-5] The intermolecular energy of the crystal (the lattice energy) is approximated by a sum of pairwise *inter*-molecular atom-atom interactions with the molecules treated as conformationally rigid (thereby neglecting *intra*-molecular interactions). The crystal forcefield therefore consists of only the van der Waals and electrostatic terms from Equation 3.1, summed over the intermolecular interactions:-

$$E = \sum \varepsilon_{ij}\left[\left(\frac{r^{*}_{ij}}{r_{ij}}\right)^{12} - 2\left(\frac{r^{*}_{ij}}{r_{ij}}\right)^{6}\right] + \sum \frac{q_i q_j}{r_{ij}}$$

Equation 5.1

Note that there is no specific term to account for hydrogen bonding, which is considered by this forcefield to be a wholly electrostatic interaction. That this was sufficient for reproducing hydrogen bonded crystal structures was one of the major findings of the work by Lifson[2,3] and is in broad agreement with the results presented in this chapter. The essentially electrostatic nature of the hydrogen bond is not a new concept and was emphasised by Coulson and Danielsson as early as 1954.[7]

The values of the nonbond parameters r^{*}_{ii} and ε_{ii} are given in Table 5.1. These parameters are used for interactions between atoms of the same type. The

combination rule used for two different types of atoms i and j is

$$r^*_{ij} = \tfrac{1}{2}(r^*_{ii} + r^*_{jj}) \quad \text{and} \quad \varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$$

Note that the r^*_{ii} and ε_{ii} parameters for the hydroxyl hydrogens are set to zero in Table 5.1. These values were chosen by Lifson's group because optimisations of these parameters to carboxylic acid crystals gave such large standard deviations that no meaningful values could be assigned to them.[3] The justification made for ignoring the van der Waals interactions of these hydrogens is that, in calculations of electronic distributions of **X–H** diatomic molecules, the 'size' of the hydrogen decreases significantly with the electronegativity of **X**. For highly electronegative **X** atoms (like oxygen) the electronic distribution is approximately spherical around **X** - and the hydrogen is now so small that it can be neglected.[8]

Table 5.1 Non-Bond Parameters[a]

Atom Type	r^*_{ii}	ε_{ii}	q_i
C	4.35	0.039	See text
O	3.21	0.228	-0.38
H(-C)	2.75	0.038	+0.10
H(-O)	0	0	+0.35

a r^*_{ii} values are in Å, ε_{ii} values in kcal/mol.

Partial charge values (q_i) for the **O**, **H**(-C) and **H**(-O) atoms were again taken from the amide and acid optimisations. Charges on carbon atoms were selected to give electroneutrality by balancing the charges on their substituent atoms. Thus the partial charge value (q_C) for a given carbon is determined by

$$q_C = 0.19\,n_O + 0.03\,n_{OH} - 0.10\,n_H$$

where n_O, n_{OH} and n_H are the number of ether oxygens, hydroxyl groups and hydrogens bonded to the carbon atom respectively. This is a similar approach to the one used by the Lifson group, who assumed electronegativity of the constituent functional groups of their model compounds.

Although not directly related to the crystal simulations, a discussion on dipole moments has been included at the end of this chapter (Section 5.5). Dipole moments are a direct indicator of charge distribution within the molecule and are therefore closely related to the selection of q_i values.

5.3 The Crystal Model

The crystal model is generated by considering a single unit cell (which may contain more than one molecule) in the centre of a three dimensional array of identical unit cells. The sums in Equation 5.1 are made over the intermolecular non-bond interactions between each atom in the unit cell and all other atoms in the crystal model within a specified cut-off range.

The greater the cut-off range, the more time-consuming will be the calculation, as a larger number of pairwise interactions will have to be computed. On the other hand, too short a cut-off range may mean that some interactions longer than this distance, but nevertheless having an important contribution to the lattice energy, will be neglected. This would result in an inaccurate summation limit of the lattice energy (E in Equation 5.1). In order to maintain accuracy, the value of the cut-off was therefore chosen to be as large as possible within the constraints of practicable computational time limits.

In our calculations we used a 15Å cutoff, which Kitaigorodsky recommends as the necessary value to achieve an accuracy to within 1% for the summation limit of

the lattice energy.[9] In fact, the cut-off is applied in such a way that, if *any* atom of a molecule is within 15Å of the unit cell, the entire molecule is included. Thus the effective cut-off is significantly larger than 15Å. Typical CPU expenditure for crystal minimisations using a cut-off range of 15Å on our computer (a DEC MicroVAX II) were between 10 and 12 hours. The nine different crystal structures studied therefore required a total in the region of 100 hours of CPU time.

In crystal simulations by Rasmussen,[10] *convergence acceleration* methods are used which reduce the computational time required. These allow a smaller cut-off distance to be used while still maintaining accuracy in the summation limit. For a more exhaustive study than the one reported here, this technique may be of assistance in keeping the computational cost within reasonable limits. However, in this instance the small number of crystal minimisations carried out did not justify the programming effort required to implement it in our program.

The total intermolecular energy of the lattice, given by Equation 5.1, is minimised with respect to the three rotational and three translational coordinates of each molecule in the unit cell, as well as the nine Cartesian components of the three unit cell vectors. The total number of variables is therefore 6n+9, where n is the number of molecules per unit cell. This is reduced to 6(n-1)+9 variables by fixing one of the molecules in space and allowing the others to move relative to it; thus removing the six rotational and translational degrees of freedom of the lattice as a whole.

In each iteration of the minimisation, the forces acting on each molecule of the central unit cell are computed and the molecules moved accordingly. The new coordinates of the unit cell are then used as a template to generate the new three dimensional array of unit cells to be used in the next iteration.

The minimisation algorithm implemented in the crystal minimisation program is of the quasi-Newton type described in Chapter 3.

5.4 Crystal Simulations of Model Compounds

Carbohydrate molecules can be considered as a combination of methylene groups, ether oxygens and hydroxyl groups. The model compounds selected for crystal simulation were therefore chosen to reflect the non-bond characteristics of these individual structural units:

Alkanes	Cyclohexane,[11] *n*-Octane[12]
Ethers	Diethylether,[13] 1,4-Dioxane (phases I & II),[14] Trioxane[15]
Alcohols	Ethanol[16]
Carbohydrates	α-D-Glucose,[17] β-D-Glucose[18]

These crystals exhibit a range of crystal packing forces, extending from hydrocarbon-hydrocarbon interactions through to crystals with extensive hydrogen bonding. The two glucose crystals were included as representative carbohydrate crystals.

The initial unit cell geometries were taken from x-ray structure determinations in all cases except α-D-glucose, which was taken from a neutron diffraction study. The unit cell dimensions, space group and coordinates of the asymmetric units for these crystals were obtained from the Cambridge Crystallographic Database. The coordinates of the complete unit cells were then generated using the interactive computer graphics package INSIGHT.[19] This package has a facility for constructing the unit cell from the asymmetric unit using the symmetry operations of the space group.

In the case of the structures found by x-ray diffraction, the locations of the hydrogen atoms (which are inaccurately determined by this method) were calculated by minimising the intramolecular energies with respect to the hydrogen coordinates while keeping all other atoms fixed at their crystallographic positions. This was performed using the DISCOVER molecular modelling package.[6] For the trioxane crystal,

where no hydrogen coordinates were given in the experimental structure, the positions of the hydrogens first had to be built assuming a standard 'tetrahedral' geometry about the carbon atoms. The hydrogen positions were then allowed to relax in the same manner.

The resulting atomic coordinates of the unit cells were used to construct the crystal lattice model, which was then minimised using the method described in the preceding section (5.3).

5.4.1 Minimised Crystal Structures

Disregarding thermal motions of the atoms, an experimental crystal structure relates to the true minimum energy geometry of the crystal. Because the experimental structure is used as the initial geometry for the calculation, a measure of the performance of the forcefield is therefore how little the crystal model has changed on minimisation.

The success of the potential energy parameters was judged by their ability to reproduce a variety of structural properties. These were the unit cell parameters (a,b,c,α,β,γ), unit cell volume, and short-range interatomic distances. Hydrogen bond lengths were also considered for those molecules containing hydroxyl groups.

The experimental and calculated unit cell parameters for the nine crystal structures studied are shown in Table 5.2. The table also shows the deviation of the unit cell parameters from those experimentally observed. We have chosen to express the deviations of the unit vector lengths (a,b and c) in terms of a percentage; the reason being that the unit cells vary a great deal in size across the nine crystals, and therefore absolute deviations (in Å) do not give a good indication of the *relative* performance of the model from one crystal to another.

Deviations in unit cell vector lengths can be seen from Table 5.2 to vary from 0.7% to -5.6% (with an average of 2.5%). Similarly, the unit cell angles show deviations ranging from -0.8° to -8.3° (with an average of 2.3°).

Table 5.2 Comparison of Experimental and Calculated Unit Cell Parameters[a]

Crystal	n[b]		a	b	c	α	β	γ
Cyclohexane		Expt.	11.23	6.44	8.20	90	108.8	90
	4	Calc.	11.43	6.39	8.34	90	107.7	90
(*C2/c*)		Dev.	+1.8%	-0.8%	+1.7%	-	-1.1	-
***n*-Octane**		Expt.	4.22	4.79	11.20	94.7	84.3	105.8
	1	Calc.	4.18	4.52	10.98	92.9	83.5	104.4
(*p 1*)		Dev.	-0.9%	-5.6%	-2.0%	-1.8	-0.8	-1.4
Diethylether		Expt.	11.81	8.07	10.85	90	90	90
	8	Calc.	12.13	8.13	10.42	90	90	90
($P2_12_12_1$)		Dev.	+2.7%	+0.7%	-3.9%	-	-	-
1,4-Dioxane		Expt.	4.58	9.18	5.82	90	99.6	90
phase I	2	Calc.	4.33	8.70	6.09	90	91.3	90
($P2_1/n$)		Dev.	-5.5%	-5.2%	+4.6%	-	-8.3	-
1,4-Dioxane		Expt.	5.74	6.51	6.14	90	100.2	90
phase II	2	Calc.	5.78	6.31	6.48	90	102.4	90
($P2_1/n$)		Dev.	+0.7%	-3.1%	+5.5%	-	+2.2	-
Trioxane		Expt.	9.32	9.32	8.20	90	90	120
	6	Calc.	9.57	9.48	8.26	92.0	90.0	120.3
(*R3c*)		Dev.	+2.7%	+1.7%	+0.7%	+2.0	0.0	+0.3
Ethanol		Expt.	5.38	6.88	8.26	90	102.2	90
	4	Calc.	5.27	6.78	8.42	90	101.4	90
(*P c*)		Dev.	-2.0%	-1.5%	+1.9%	-	-0.8	-
α-Glucose		Expt.	10.37	14.85	4.97	90	90	90
	4	Calc.	10.28	14.53	4.95	90	90	90
($P2_12_12_1$)		Dev.	-0.9%	-1.8%	-0.4%	-	-	-
β-Glucose		Expt.	9.20	12.64	6.65	90	90	90
	4	Calc.	9.63	12.36	6.41	90	90	90
($P2_12_12_1$)		Dev.	+4.7%	-2.2%	-3.6%	-	-	-

a Distances in Å, Angles in degrees
b n = Number of molecules per unit cell

The experimental errors arising from an x-ray structure determination are much smaller than these values. Estimated standard deviation values for the unit cell dimensions are quoted for each experimental structure and these equate to experimental errors of only a fraction of a percent. Another measure of the accuracy of a crystal structure is the discrepancy index (or R factor), which for the crystals studied varied from 0.028 to 0.110. A low value of R indicates an accurately determined structure, and values below 0.1 are considered good by present standards.[20]

There were no symmetry constraints imposed during the minimisation process, although the symmetry of the crystal space group was present in the initial lattice geometry. In all cases except one (trioxane), this symmetry has been maintained in the minimised structures. Theoretically, any existing symmetry *should* be preserved: minimisation operates by moving the molecules in accordance with the forces acting on them ($\partial E/\partial x_i$). Since these forces will also display the symmetry of the lattice; overall crystal symmetry will be maintained.

Why the trioxane crystal should lose its symmetry is a more difficult question. The initial crystal structure is hexagonal ($a=b\neq c$, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$) but becomes slightly distorted on minimisation. It is likely that the problem is related to the generation of the unit cell coordinates from the asymmetric unit. The symmetry operators for the space group of Trioxane (R3c) involve transformations containing a recurring decimal number in the transformation matrix (i.e. translation by $^1/_3$ of a unit cell). Therefore truncation errors may have been introduced into the coordinates which would have resulted in a loss of symmetry in the initial structure. Once symmetry has been lost, minimisation can then proceed to an asymmetric minimum. For all the other crystal structures, no such transformation matrices containing recurring decimal numbers were required. Hagler *et al.*[5] encountered similar problems of loss of symmetry when minimising the crystal structure of butanoic (or butyric) acid.

Table 5.3 Comparison of Experimental and Calculated Unit Cell Volumes and Average Change in Interatomic Distances ($|\Delta d|$) on minimisation.

Crystal	Unit Cell Volume ($Å^3$)			$\|\Delta d\|$
	Expt.	Calc.	ΔV	(Å)
Cyclohexane	561	580	+3.4%	0.11
n-Octane	216	199	-7.8%	0.60
Diethylether	1034	1027	-0.7%	0.18
1,4-Dioxane (phase I)	241	230	-4.6%	0.29
1,4-Dioxane (phase II)	226	231	+2.2%	0.14
Trioxane	617	647	+5.0%	0.19
Ethanol	299	295	-1.3%	0.12
α-D-Glucose	766	739	-3.5%	0.14
β-D-Glucose	774	764	-1.3%	0.14

Table 5.3 shows the change in unit cell volume (ΔV) on minimisation. The unit cell volume is merely a function of the unit cell parameters and does not provide any additional information as such. It does, however, provide a rough indication of how much the crystal has expanded or contracted on minimisation. Again, deviation is expressed as a percentage of the experimentally observed volume. Errors in the calculated structure range from -0.7% to -7.8% in the worst case (*n*-octane). The overall average deviation in unit cell volume is 3.3%.

Another quantity $|\Delta d|$, is given in Table 5.3. This represents the mean of the absolute differences, $|r_{exptl} - r_{calcd}|$, between the interatomic distances of less than 4Å in the experimental structure. Its units are in Å and a small value represents a good fit to experiment. The $|\Delta d|$ values in Table 5.3 are all much larger than might be

explained by experimental error. Generally, in crystal structure determinations for molecules of this size, interatomic distances are accurate to within 0.01Å.[20] The worst agreement is again in the case of *n*-octane which shows a $|\Delta d|$ value of 0.6Å. It is interesting to note that the best agreement is for cyclohexane, the other hydrocarbon crystal tested. Why this disparity should occur is uncertain, as both crystals display the same types of non-bond interactions (ie. **C····C, H····H, C····H**).

For the two carbohydrate crystals however (the class of compounds for which this forcefield is ultimately aimed) the $|\Delta d|$ values are encouragingly small (0.14Å in both cases).

A comparison of these calculations with the previous ones by Hagler, Dauber and Lifson (HD&L) is made in Table 5.4. The figures shown for the HD&L calculations are based on the unit cell parameters for the 14 carboxylic acid crystals given by the '12-6-1 potential' in reference 5. The quantities $\Delta(a,b,c)$ and $\Delta(\alpha,\beta,\gamma)$ are the deviations in unit cell vector lengths and intersection angles respectively. As can be seen from Table 5.4, the calculations from this work are shown to compare favourably, having similarly low deviations from experimental structure to those of the carboxylic acid simulations.

Hydrogen bonding deserves special attention because it is a major feature of carbohydrate non-bonded interactions. Table 5.5 shows the deviation in hydrogen bond lengths for the three crystals studied in which they occur. The values given in the table relate to the **O····O** distances rather than the **H····O** distances, because, as discussed above, the hydrogen positions are often poorly determined by x-ray. Two different length hydrogen bonds occur in the ethanol crystal, due to the fact that both the *gauche* and *trans* conformations of ethanol are present. For α-D-glucose and β-D-glucose, it can be seen that there are respectively 5 and 4 intermolecular hydrogen bonds per molecule. In all cases the hydrogen bond distance is shorter in the calculated structure than the experimental. This may be due to the neglect of the

Table 5.4 Comparison of this work with previous calculations (H,D & L)[5]

		Δ(a,b,c)	Δ(α,β,γ)	ΔV
H,D & L	Maximum	20.3%	6.2°	-8.6%
	Minimum	-0.2%	0.2°	-1.2%
	Average	4.3%	2.0°	3.8%
This Work	Maximum	-5.6%	-8.3°	-7.8%
	Minimum	0.7%	-0.8°	-0.7%
	Average	2.5%	2.3°	3.3%

Table 5.5 Interatomic Distances between Oxygen Atoms linked by Hydrogen Bonds (Å)

Crystal	Expt.	Calc.	Devn.[a]
Ethanol	2.72	2.66	-0.06
	2.73	2.69	-0.04
α-D-Glucose	2.71	2.60	-0.11
	2.71	2.62	-0.10
	2.78	2.69	-0.08
	2.85	2.69	-0.16
	2.78	2.69	-0.09
β-D-Glucose	2.71	2.59	-0.12
	2.68	2.67	-0.02
	2.77	2.69	-0.08
	2.67	2.59	-0.08

a Devn = Calc - Expt

van der Waals effects of the hydroxyl hydrogen, or alternatively the q_i parameters of **O** or **H** may be slightly in error. The average shortening is by 0.085Å, and this effect is probably responsible for the slight contraction seen in the unit cell volume of these crystals.

5.4.2 Crystal Lattice Energies

In addition to the structure of the crystals studied, the thermodynamic properties are also of interest. The crystal forcefield equation (5.1) gives the total lattice energy as a sum of a van der Waals and an electrostatic term. The lattice energies (E_{Tot}) of the initial and minimised crystal structures together with their van der Waals (E_{vdW}) and electrostatic (E_{elec}) components are given in Table 5.6. Note that E_L, the lattice energy per mole, is obtained by dividing E_{Tot} by the number of molecules per unit cell.

The lattice energies E_{Tot} do not change much on minimisation (generally <4 kcal/mol) even in the cases where the change in unit cell dimensions is relatively large. This 'shallowness' of the crystal energy surface was also found for both carboxylic acid and amide crystals.[4,5] The relative contributions of the electrostatic interactions to the total energy can be seen to increase as the number of ether oxygens and hydroxyl groups per molecule increase. The lattice energy for cyclohexane, for example, is almost exclusively the result of van der Waals interactions (>99%); while in the glucose structures the electrostatic interactions (including hydrogen bonding) become the major contributor (>60% of E_{Tot}). This is in agreement with what we might intuitively expect when considering the packing forces in lattices of these molecules.

From the lattice energies per mole (E_L), some conclusions can also be drawn about the relative stabilities of the lattices. The two crystal phases of 1,4-Dioxane are shown to differ in terms of intermolecular energy energy by 1.13 kcal/mol, with phase I being the more stable. Experimentally, phase II is found to exist at lower

Table 5.6 Lattice Energies of the Calculated Structures (kcal/mol)

Crystal	n[a]	Initial Structure			Final Structure[b]			
		E_{elec}	E_{vdW}	E_{Tot}	E_{elec}	E_{vdW}	E_{Tot}	E_L
Cyclohexane	4	-0.60	-46.96	-47.56	-0.42 (0.9%)	-48.29 (99.1%)	-48.71	-12.18
***n*-Octane**	1	1.05	-16.93	-15.88	1.10 (6%)	-17.93 (94%)	-16.82	-16.82
Diethylether	8	-13.63	-84.27	-97.90	-15.61 (16%)	-84.75 (84%)	-100.36	-12.55
1,4-Dioxane (phase I)	2	-8.75	-23.05	-31.80	-9.12 (27%)	-24.92 (73%)	-34.04	-17.02
1,4-Dioxane (phase II)	2	-7.75	-23.29	-31.04	-7.22 (23%)	-24.56 (77%)	-31.78	-15.89
Trioxane	6	-21.44	-66.06	-87.50	-19.83 (22%)	-71.47 (78%)	-91.30	-15.22
Ethanol	4	-15.51	-24.53	-40.04	-17.97 (44%)	-23.16 (56%)	-41.13	-10.28
α-D-Glucose	4	-94.90	-68.05	-162.95	-109.73 (64%)	-61.60 (36%)	-171.32	-42.83
β-D-Glucose	4	-82.84	-66.22	-149.06	-92.89 (60%)	-61.01 (40%)	-153.90	-38.48

a n = Number of molecules per unit cell

b Percentage values show relative contributions of E_{elec} and E_{vdW} to E_{Tot}

temperatures than phase I[14] indicating the reverse to be the case. These calculations take no account of intramolecular strain energy of course, which may stabilise phase II with respect to phase I.

α-D-Glucose can also be seen from the table to have a lower lattice energy than the β- form by 4.35 kcal/mol. This is probably due to the additional intermolecular hydrogen bond observed in the crystal structure of the α- form (see Table 5.5).

5.4.3 Sublimation Energies

Some comparison with experiment may be made for calculated lattice energies by considering enthalpies of sublimation of the crystals. Unfortunately, however, the limited availability and accuracy of this type of thermodynamic data do not allow such a comprehensive and reliable comparison with experiment as was made for the structural properties of the crystals.

A theoretical value for the sublimation energy (ΔH_s) may be calculated from the lattice energy per mole (E_L) as follows :-

$$\Delta H_s = H_{gas} - H_{solid}$$

$$\approx pv + 3RT - (E_L + 6RT)$$

$$\approx -E_L - 2RT$$

Here we have assumed H_{gas} to be given by the ideal gas law and the vibrational energy of the crystal to be 6RT. At 298 K the value of 2RT is 1.18 kcal/mol, thus:-

$$\Delta H_{s\,(calc)} = -E_L - 1.18$$

In the transition from the solid to the gas phase both geometrical and vibrational changes occur because of the removal of crystal packing forces on the molecules. This would also make a contribution to the sublimation energy, but since this contribution is likely to be relatively small (0-2 kcal/mol)[21] and little information regarding these effects is available, we have chosen to ignore them in our estimations. In the case of cyclohexane and trioxane, experimentally determined heats of sublimation were available for comparison.[22,23] For some of the other crystals, the sublimation energy was estimated by adding the heats of melting (ΔH_m) and vaporisation (ΔH_v) adjusted to 298 K:-

$$\Delta H_{s\,(expt)} = \Delta H_m + \Delta H_v$$

ΔH_m and ΔH_v values were taken from the 69th C.R.C Handbook.[22]

A comparison of these calculated and 'experimental' heats of sublimation is made in Table 5.7. For the glucose crystals, neither heats of melting or vaporisation were available, so no comparison of the calculated ΔH_s values can be made.

As can be seen from the table, the calculated and experimental values differ by as much as 4.2 kcal/mol. This is perhaps unsurprising when one considers the approximations made in calculating ΔH_s(calc.), and the estimation of ΔH_s(expt.) from ΔH_m and ΔH_v values. Experimental errors of the ΔH_s, ΔH_m and ΔH_v values may also be partly responsible for these discrepancies. Thermodynamic quantities of this size are very difficult to measure with accuracy. This is illustrated by the different literature values for ΔH_s of Trioxane, which has been determined as 11.60 ± 0.6 kcal/mol and 13.50 ± 0.02 kcal/mol in two separate studies.[22,24]

Table 5.7 Heats of Sublimation (kcal/mol)

Crystal	ΔH_s(calc.)	ΔH_s(expt.)	Difference	Source of ΔH_s(expt.)[a]
Cyclohexane	11.0	8.9	2.1	Measured ΔHs
n-Octane	15.6	14.2	1.4	$\Delta H_m + \Delta H_v$
Diethylether	11.4	8.7	2.7	$\Delta H_m + \Delta H_v$
1,4-Dioxane (phase I)	15.8	11.6	4.2	$\Delta H_m + \Delta H_v$
1,4-Dioxane (phase II)	14.7	-	-	-
Trioxane	14.0	13.1	0.9	Measured ΔHs
Ethanol	9.1	10.0	-0.9	$\Delta H_m + \Delta H_v$
α-D-Glucose	41.7	-	-	-
β-D-Glucose	37.3	-	-	-

a See text for appropriate references

5.5 Dipole Moments

Dipole moments have been calculated from the partial charge parameters in order to compare them with experimental values. The dipole moment (μ) is a vector quantity related to the distribution of charge within the molecule, and can be calculated from the following summation

$$\mu = \Sigma \mathbf{x}_i q_i$$

where $\mathbf{x}_i$ is the vector denoting the position of atom i and q_i is its partial charge.

Determinations of dipole moments can be made in the gas phase, liquid phase and solution, but only gas phase measurements were used for comparison because they relate more closely to calculations on an isolated molecule.

The dipole moment of a molecule, being a function of the atomic positions, will be dependent on its conformation. Molecules with only a limited number of conformations were selected for comparison, thereby leaving less uncertainty as to which conformation the experimental dipole moment relates.

In general, calculated dipole moments are not particularly accurate as calculational approaches are seldom chosen to give good dipole moments.[27] It can be seen from Table 5.8 that our forcefield also has limited success. For the hydrocarbons, the dipole moments, which are experimentally found to be very small, are even smaller by calculation. Ether molecules give the worst results, with dipole moments this time *overestimated* by between 0.33 and 0.69 debyes. The alcohols in Table 5.8 can be seen to give the closest agreement with experiment, with most errors less than 0.1 debye.

5.6 Summary of Crystal Simulations

Although the parameters for the oxygen atom were originally derived for carboxylic acid and amide crystals, the non-bond part of the forcefield is found to give a reasonable account of both the structural and thermodynamic properties of the nine crystals modelled. In particular, the properties of the three crystals containing hydrogen bonds (ethanol, α- and β-D-glucose) were reproduced well even though the forcefield contained no specific term to account for hydrogen bonding. The non-bond parameters have therefore been used throughout this work without further adjustment.

Table 5.8 Comparison of Calculated and Experimental Dipole Moments (Debyes).[a]

Compound	μ_{expt}	μ_{calc}	$\mu_{calc} - \mu_{expt}$
Propane	0.084	0.010	-0.074
i-Butane	0.132	0.017	-0.115
Dimethylether	1.30	1.990	0.69
Tetrahydrofuran	1.63	2.029	0.40
Tetrahydropyran[b]	1.600	1.929	0.33
1,3-Dioxane[c]	2.14	2.675	0.54
Trioxane	2.08	2.916	0.84
Methanol	1.700	1.579	-0.121
Ethanol[d]	1.69	1.579/1.613	-0.11/-0.08
i-Propanol[d]	1.66	1.594/1.622	-0.07/-0.04
t-Butanol	1.640	1.606	-0.034

a Values taken from reference 22 unless otherwise stated.

b See reference 25

c See reference 26

d Two possible conformers exist for these molecules. Calculated values are for *trans/gauche* conformers respectively.

5.7 References to Chapter 5

1. A. Warshel and S. Lifson, *J. Chem. Phys.*, **53**, 582 (1970).
2. A. T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
3. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979).
4. A. T. Hagler, S. Lifson, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5122 (1979).
5. A. T. Hagler, P. Dauber, and S. Lifson, *J. Am. Chem. Soc.*, **101**, 5131 (1979).
6. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).
7. C. A. Coulson and U. Danielsson, *Arkiv for Fysik*, **8**, 239-255 (1954).
8. R.W.F. Bader, I. Keavany, and P.E. Cade, *J. Chem. Phys.*, **47**, 3383 (1967).
9. A. I. Kitaigorodsky, *Molecular Crystals and Molecules,* Academic Press, New York (1973).
10. K. Rasmussen, *Potential Energy Functions in Conformational Analysis,* in *Lecture Notes in Chemistry, Vol 37*, Springer-Verlag, Berlin & Heidelberg (1985).
11. R. Kahn, R. Fourme, D. Andre, and M. Renaud, *Acta Cryst. Ser. B*, **29**, 131 (1973).
12. H. Mathisen, N. Norman, and B.F. Pedersen, *Acta Chem Scand.*, **21**, 127 (1967).
13. D. Andre, R. Fourme, and K. Zechmeister, *Acta Cryst. Ser. B*, **28**, 2389 (1972).
14. J. Buschmann, E. Muller, and P. Luger, *Acta Cryst. Ser. C*, **42**, 873 (1986).
15. V. Busetti, A. Del Pra, and M. Mammi, *Acta Cryst. Ser. B*, **25**, 1191 (1969).
16. P.-G. Jonsson, *Acta Cryst. Ser. B*, **32**, 232 (1976).
17. G.M. Brown and H.A. Levy, *Acta Cryst. Ser. B*, **35**, 656 (1979).
18. S.S.C. Chu and G.A. Jeffrey, *Acta Cryst. Ser. B*, **24**, 830 (1968).

19. Available from Biosym Technologies Inc., San Diego, California, USA

20. J. Pickworth Glusker and K.N. Trueblood, in *Crystal Structure Analysis: A Primer*, Oxford University Press, New York (1972).

21. J. L. Derissen, *J. Mol. Struct.*, **38**, 177 (1977).

22. *CRC Handbook of Chemistry and Physics, 69th Edition,* The Chemical Rubber Co., Cleveland, Ohio (1988-1989).

23. H.G.M. de Wit, J.C. van Miltenberg, and C.G. de Kruif, *J. Chem. Thermodyn.*, **15(7)**, 651-3 (1983).

24. J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometalic Compounds,* Academic Press, New York, NY (1970).

25. H.E. Breed, G. Gundersen, and R. Seip, *Acta Chem. Scand.*, **A33**, 225 (1979).

26. G. Schultz and J. Hargitai, *Acta Chim. Acad. Sci. Hungary*, **83**, 331 (1971).

27. S. Melberg and K. Rasmussen, *J. Mol. Struc.*, **57**, 215 (1979).

Chapter 6

Application of the Forcefield : Results for Model Compounds

6.1 Introduction

This chapter documents the results obtained by the forcefield when applied to a selection of model compounds. The final parameter set used for all the calculations reported in this chapter is given in Appendix I. The results presented are arranged according to the types of physical property studied. Thus, molecular geometries, vibrational frequencies, rotameric energies and conformational energies are each discussed in separate sections.

Although a full 'benchmark' comparison with other forcefields has not been made here, some comparisons have been drawn between our results and those of the forcefield developed by Rasmussen and co-workers,[1-3] which is the most thoroughly documented forcefield derived for carbohydrates in the literature.

One class of compound that is absent from this chapter is acetals. These compounds exhibit the *anomeric effect*, which necessitated the introduction of a new cross-term into the potential energy function. As this formed a large area of study in its own right, a discussion of the treatment of the anomeric effect, and the results obtained for acetals, is left until the next chapter.

6.2 Molecular Geometries

Calculated molecular geometries were obtained by minimisation of initial estimated geometries using the minimisation techniques described in Chapter 3. An idealised initial geometry was chosen for all the molecules studied, with valence angles set to

the tetrahedral values of 109.5° and bond lengths set to the following standard values: 1.53 Å (**C–C**), 1.43 Å (**C–O**), 1.10 Å (**C–H**) and 0.96 Å (**O–H**). Torsion angles were set in most cases to either the *gauche* (±60°) or *trans* (180°) values. In the case of five-membered rings, however, which were initially constructed as planar, torsion angles were assigned values of 0° or ±120° as appropriate.

In general, the b_0 and θ_0 parameters were fitted to experimental molecular geometries, while the force constant parameters were fitted to vibrational frequencies. Ten molecules were used in the optimisation of the b_0 and θ_0 parameters; these were: ethane, propane, *n*-butane, *i*-butane, cyclohexane, dimethylether, ethyl-methylether, 1,4-dioxane, methanol and ethanol.

The minimised molecular geometries (bond lengths, valence angles and torsion angles) are given in Tables 6.1 to 6.3, and compared with experimental values. Gas phase structural measurements were available for all the molecules studied, although they are a mixture of r_α°, r_g, and r_a values (from electron diffraction studies) and r_0 and r_s values (from microwave determinations). A discussion of these different definitions of molecular structure was made in Chapter 2.

Some points common to all molecular geometries are pertinent at this stage, before looking in more detail at the individual compounds. For all molecules studied, the observed symmetry was reproduced in the calculated geometries, even if it was not present in the initial structure.

For the most part, the calculated geometries are in good agreement with the experiment, showing deviations within our criteria for an acceptable fit (outlined in Section 4.4). In general, the largest deviations are seen to occur in **C–H** bond lengths, with a maximum of 0.013 Å. These deviations may be due, at least in part, to experimental uncertainty: it is known that bond lengths and angles involving hydrogen atoms are strongly influenced by rotation-vibration effects and are often imprecisely determined,[4] hence uncertainties in **C–H** and **O–H** bond lengths can be several times

larger than those for **C–C** or **C–O** bond lengths.[5-7]

It should also be born in mind that structures derived from experimental results are often solved subject to assumptions as to symmetry or fixed values of certain internal coordinates. This, together with the different types of experimental structures used (r_g, r_a, r_o etc.) may also contribute to some of the larger discrepancies.

6.2.1 Hydrocarbons

Calculated and experimental hydrocarbon geometries are shown in Table 6.1. For all the hydrocarbons studied, electron diffraction data was available.

For ethane, an accurate r_α^o structure has been determined.[8] The **C–C** bond length can be seen to be slightly too short by calculation, but the valence angles are reproduced well.

The next three molecules in the table, propane, *i*-butane and neopentane differ by the addition of successive methyl groups to the central carbon atom. The increasing steric crowding in these molecules can be seen experimentally to result in an increase in the **C–C** lengths, and this is reproduced well by the forcefield. In *i*-butane, although the **C–H** bond lengths are in error, the longer methine **C–H** is predicted by calculation. The **C–C–C** bond angle in neopentane is constrained, by reasons of symmetry, to adopt the tetrahedral value (109.5°) in both the experimental and calculated structures.

n-Butane is of particular interest because it contains a **C–C–C–C** torsion that is acyclic, and so not constrained by being in a ring. The calculated value for the *gauche* torsion angle (68°) matches that found experimentally (65°) very well indeed. Both the *gauche* and *trans* rotamers were minimised, as the experimental structure corresponds to an average geometry of a mixture of the two rotamers in the gas phase. The experimental structure was determined with the following two assumptions: firstly, that all three **C–C** bond lengths were identical, and secondly that the

Table 6.1 Comparison of Experimental and Calculated Hydrocarbon Molecular Geometries[a]

Compound	Internal[b]	Exptl.[c,d]	Calc.[d]	Diff.[e]	Reference
Ethane	**C–C**	1.532	1.527	-0.005	r_α^0 (8)
	C–H	1.102	1.106	0.004	
	H–C–C	111.4	111.7	0.3	
	H–C–H	107.5	107.1	-0.4	
Propane	**C–H**	*1.096*	*1.107*	*0.011*	r_α^0 (9)
	C–C	1.531	1.532	0.001	
	C–C–C	112.4	113.8	1.4	
i-Butane	**C_t–H**	1.122	1.109	-0.013	r_α^0 (10)
	C_m–H	*1.113*	*1.106*	*-0.007*	
	C–C	1.535	1.537	0.002	
	C–C–C	110.8	111.4	0.6	
Neopentane	**C–H**	1.120	1.107	-0.013	r_g (11)
	C–C	1.539	1.540	0.001	
	C–C–C	109.5	109.5	0	
n-Butane			*trans/gauche*		r_g (7)
	C–H	*1.117*	*1.107*	*-0.010*	
	C_s–C_s	*1.531*	1.542/1.545	0.011/0.014	
	C_s–C_m		1.535/1.536	0.004/0.005	
	C–C–C	113.8	114.3/116.3	0.5/2.5	
	C–C–C–C	180/65	180/68	0/3	
tri-(*t*-Butyl)-methane	**C_t–C_q**	1.611	1.681	0.070	r_g (12)
	C_q–C_m	*1.548*	*1.555*	*0.007*	
	C–H	*1.111*	*1.105*	*-0.006*	
	H–C_t–C_q	101.6	100.7	-0.9	
	C_q–C_t–$C_{q'}$	116.0	116.7	0.7	
	C_t–C_q–C_m	*113.0*	*114.6*	*1.6*	
	C_m–C_q–$C_{m'}$	*105.8*	*103.9*	*-1.9*	
Cyclohexane	**C–H**	*1.103*	*1.107*	*-0.004*	r_α (6)
	C–C	1.531	1.532	0.001	
	H–C–H	107.5	105.4	-2.1	
	C–C–C	111.4	113.1	1.7	
	C–C–C–C	±55	±50	-5	
Cyclopentane	**C–H**	*1.095*	*1.108*	*0.013*	r_a (13)
	C–C	*1.539*	*1.530*	*-0.009*	

a Bond lengths are in Å, Bond angles in degrees.

b Carbon atom subscripts: m = methyl, s = secondary, t = tertiary, q = quaternary.

c Values are from gas phase electron diffraction studies.

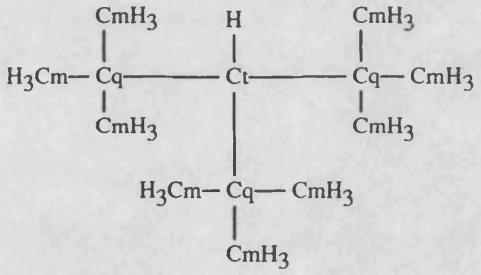
d values in *italics* indicate the internals which are assumed in the experimental model to be equivalent throughout the molecule (e.g. *all* **C–H** lengths in Propane were assumed to be equal[9]). The appropriate calculated values are averaged to facilitate comparison.

e Diff = Calc - Exptl

gauche and *trans* forms differed only in the value of the torsion angle. The calculated **C–C** bonds are rather long (average: 1.537 Å *trans*; 1.539 Å *gauche*) and contrary to the experimental assumptions *are* different, with the central bond longer than the terminal ones. This is in accord with high level basis set *ab initio* calculations, which also indicate a longer central **C–C** bond in *n*-butane.[14]

Note that our calculated values for *n*-butane also suggest a stretching of the central bond on conversion from the *trans* to the *gauche* rotamer, together with an opening of the **C–C–C** angles. This occurs due to an increased repulsion between the two methyl groups in the *gauche* form.

Tri-(*t*-butyl)-methane is an extremely crowded molecule that has been used as a test case for other forcefields.[3,15] It was not used by us in the derivation of the parameters, but we have included it here to examine the applicability of the forcefield to highly strained molecules.



t = tertiary
q = quaternary
m = methyl

tri-(t-Butyl)-methane

Although the structure of tri-(*t*-butyl)-methane is for the most part reasonably well reproduced, the calculated $\mathbf{C_t–C_q}$ bonds are considerably too long (by 0.07 Å). Other workers,[3,15] have achieved better results for this bond length, but this is probably due to the absence of electrostatic terms[15] (the main source of the non-bond repulsion in our calculations) or the use of a harmonic bond potential[3] (which is steeper than the Morse potential at bond lengths far from b_0).

The last two molecules in Table 6.1, cyclohexane and cyclopentane, are model compounds for molecules containing six- and five-membered rings. For cyclohexane, calculated values are in good agreement with experiment. However, this required another atom type (C_6) to be specified in the forcefield to account for carbon atoms in six-membered rings, as the **C–C** bond lengths were too long (by 0.02 Å) using normal carbon atom parameters. The C_6 atom type differs from a standard carbon atom in that it has a slightly shorter b_0 value for its endocyclic bonds (see Appendix I). Otherwise it is identical to a standard carbon atom (**C**).

The reason for the over-stretching of the bonds in cyclohexane was due to the large non-bond repulsions (van der Waals and electrostatic) that occur between opposite atoms in the ring. Non-bond effects are considered by the forcefield to operate between atoms 1,4 to each other (separated by three bonds) or further. There are three such 1,4 interactions across a cyclohexane ring (between the three pairs of opposite atoms) and the stretching force on the **C–C** bonds is therefore very large. The choice of a shorter b_0 as a remedy for this problem does not seem to have affected the calculated vibrations adversely, which are shown in Section 6.3.1 to be in excellent agreement with experiment.

Apart from the **C–C** bond lengths, the other structural features are reasonably reproduced using standard hydrocarbon parameters, although the slight overestimation of the **C–C–C** angles leads to a less puckered conformation than found by experiment.

The problem of how 1,4 interactions should be dealt with is not new. Some, like ourselves, include both non-bond terms and torsional terms to account for these interactions, while Rasmussen attempts to account for 1,4 interactions by non-bond terms alone.[1] Another course of action is taken by the authors of the AMBER forcefield,[16] which includes torsional terms but scales 1,4 non-bond interactions by half; although the reasons for choosing this scale factor seem rather arbitrary.

The only structural parameters to have been determined for cyclopentane are the **C–C** and **C–H** bond lengths, as experimentally no well-defined conformation is observed due to pseudorotation.[13] The calculated minimum is found to be the C_2 (or *twist* form) rather than the C_s (*envelope*) form, regardless of the initial conformation used. Again, we have problems in reproducing endocyclic bond lengths (a problem shared by other forcefields[3]). Contrary to cyclohexane, however, the calculated **C–C** bond lengths in cyclopentane are *too short*. This is because in cyclopentane there are no non-bond effects considered between ring atoms (as no two ring atoms are separated by more than two bonds). Although not yet implemented, when more accurate structural data is available for cyclopentane, there may be a case for the generation of another atom type in the forcefield, this time for carbon atoms in a five-membered ring, and having a *longer* b_0 value than a standard carbon atom.

6.2.2 Ethers

Minimised geometries of the ether model compounds are shown in Table 6.2. As for the hydrocarbons, all experimental data was taken from gas phase electron diffraction studies.

A general feature of all the ether molecules is that the calculated **C–O–C** angles are larger than the corresponding experimental ones. This is a general problem for forcefields that do not explicitly include lone-pair electrons, as repulsions between the lone-pair and bond-pair electrons are not accounted for. These repulsions are partly responsible for keeping the **C–O–C** angle from opening up.

In order to keep the **C–O–C** angles reasonably close to experiment, we chose a small value for the θ_0(**COC**) of 104.0°. An even smaller value could be used to give better agreement with experiment, but this would have resulted in a poor fit to the vibrational frequencies. The value of 104.0° was therefore chosen as a compromise that fits both the geometrical and vibrational data reasonably well. Other forcefields

Table 6.2 Comparison of Experimental and Calculated Ether Molecular Geometries[a]

Compound	Internal	Exptl.[b,c]	Calc.[c]	Diff.[d]	Reference
Dimethylether	**C–H**	*1.094*	*1.107*	*0.013*	r_a (5)
	C–O	1.416	1.420	0.004	
	C–O–C	111.5	114.0	2.5	
Ethylmethylether[e]			*trans/gauche*		r_α^0 (17)
	C–H	*1.118*	*1.107*	*-0.011*	
	$\mathbf{C_m}$**–O**	1.413	1.421/1.423	0.008/0.010	
	$\mathbf{C_s}$**–O**	1.422	1.419/1.425	-0.003/0.003	
	C–C	1.520	1.527/1.532	0.007/0.012	
	C–O–C	111.9	114.6/116.2	2.7/4.3	
	C–C–O	109.4	109.2/113.5	-0.2/4.1	
	C–C–O–C	180/84	180.0/83.1	0/1	
1,4-Dioxane	**C–H**	*1.112*	*1.107*	*-0.005*	r_a (18)
	C–O	1.423	1.424	0.001	
	C–C	1.523	1.530	0.007	
	C–O–C	112.5	114.2	1.7	
	C–C–O	109.2	112.5	3.3	
	O–C–C–O	±58	±49.3	-9	
	C–C–O–C	±57	±50.1	-7	
Tetrahydrofuran	**C–O**	1.426	1.421	-0.005	r_g (19)
	$\mathbf{C_2}$**–**$\mathbf{C_3}$	*1.535*	1.525	-0.010	
	$\mathbf{C_3}$**–**$\mathbf{C_4}$		1.519	-0.016	
	C–O–C	106.4-110.6	106.5	-	
	C–C–O	104.0-107.5	109.5	-	
	C–C–C	101.5-104.4	101.1	-	
	C–O–C–C	1.0-40.5	11.6	-	
	O–C–C–C	0.9-37.5	29.6	-	
	C–C–C–C	0.0-35.4	33.7	-	
Tetrahydropyran	**C–O**	1.420	1.422	0.002	r_α^0 (20)
	C–C	*1.581*	*1.580*	*-0.001*	
	C–H	*1.116*	*1.107*	*-0.009*	
	C–O–C	111.5	115.0	3.5	
	C–C–O	111.8	113.3	1.5	
	$\mathbf{C_2–C_3–C_4}$	108.3	111.7	3.4	
	$\mathbf{C_3–C_4–C_5}$	110.9	112.4	1.5	
	C–O–C–C	±59.9	±54.6	-5.3	
	O–C–C–C	±56.9	±49.9	-7.0	
	C–C–C–C	±52.5	±46.6	-5.9	

a Bond lengths are in Å, Bond angles in degrees.

b Values are from gas phase electron diffraction studies.

c Values in *italics* indicate the internals which are assumed in the experimental model to be equivalent throughout the molecule (e.g. *all* **C–H** lengths in Dimethylether were assumed to be equal[5]). The appropriate calculated values are averaged to facilitate comparison.

d Diff = Calc - Exptl

e Carbon atom subscripts: m = methyl, s = secondary (methylene).

use similarly small values of θ_0 for the **C–O–C** angle, for much the same reasons[1] and even one forcefield that does include lone-pair electrons explicitly (as dummy atoms) uses a small θ_0 value of 104.1° and a large K_θ value,[21] presumably to keep the **C–O–C** angle sufficiently closed.

For dimethylether, in addition to the **C–O–C** valence angle, the short **C–O** bond lengths are also slightly overestimated by our calculation. Two independent studies in the same year (1959) gave very similar geometries and so the experimental values are unlikely to be in error.[5,22]

It is interesting to note that, as for dimethylether, the **O–C**(methyl) bond in ethyl-methylether is found experimentally to be shorter than other **O–C**(alkyl) bond lengths. It may be that the 'non-methyl' **C–O** bonds are in fact lengthened by hyperconjugation of hydrogens 1,4 to the oxygen that cannot occur in methyl **C–O** bonds (Figure 6.1).

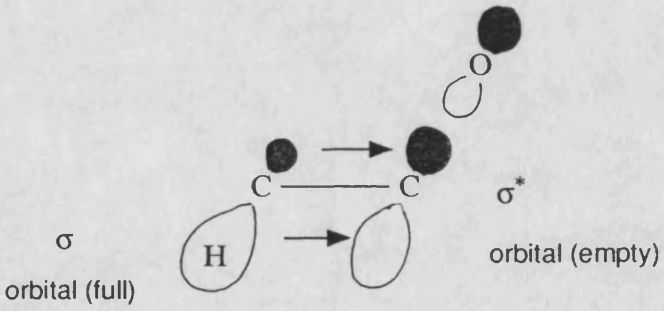


Figure 6.1

This would also explain why **C–C** bonds adjacent to **C–O** bonds are often shorter than in hydrocarbons, as the above mechanism would suggest a partial double bond between the two carbons. Similar mechanisms have been used to explain bond length changes in hydrocarbons with fluorine heteroatoms,[23] and the parallels of this mechanism with the *gauche effect* and the *anomeric effect* will become apparent later in this chapter and the next.

Although the experimental geometry of ethylmethylether was determined mainly by electron diffraction, the two different **C–O** bond lengths were in fact resolved by microwave spectroscopy, as the rotational constants of the molecule depend heavily on the relative **C–O** lengths.[17] As in the case of *n*-butane, the experimental geometry was determined by assuming identical bond lengths and valence angles for the *gauche* and *trans* rotamers. Both rotamers were minimised, and can be seen to compare reasonably well with experiment. In particular, the torsion angle of the *gauche* form is nicely reproduced (the *trans* form has a torsion value of 180° due to symmetry). Again, similarly to *n*-butane, the calculated central **C–C** bond stretches, and the backbone valence angles open, on going from the *trans* to the *gauche* rotamer. This relieves some of the repulsive interactions between the terminal methyl groups.

The experimental structure for 1,4-dioxane is taken from an old determination with fairly large experimental uncertainties (±0.005 Å in bond lengths, ±0.5° in valence angles).[18] 1,4-Dioxane also poses additional problems to the other ether compounds studied, in that it possesses vicinally disubstituted oxygen atoms. These lead to a stereoelectronic effect known as the *gauche effect* (see Section 6.4.2) which could result in distortions not seen in the other ethers. In view of these considerations, the deviations in the calculated values for 1,4-Dioxane are not too large.

The last two ether molecules in Table 6.2, tetrahydrofuran and tetrahydropyran, are model compounds for the furanose and pyranose rings so commonly found in carbohydrates. Tetrahydrofuran, like cyclopentane, is a pseudorotator and so well-defined conformations are not observed.[19] Regardless of the initial conformation, however, the only minimum energy conformation was the C_2 form (as was also the case for cyclopentane). In the structure determination of Geise *et al.*,[19] bond lengths were proposed for each of the possible conformations (C_2 and C_s). The values proposed for the C_2 form are therefore used for comparison with the calculated values. For the same reasons as those for cyclopentane, the endocyclic bond lengths are

slightly too short, but the valence and torsion angles generally fall within the ranges determined by experiment.

For tetrahydropyran, the calculated structure can be seen from Table 6.2 to reproduce the bond lengths well. The slight overestimation of the ring valence angles is undoubtably the cause of the errors in the torsion angles, as the ring becomes flatter as the valence angles increase. Overall, however, the structure of tetrahydropyran is again reasonably well reproduced.

6.2.3 Alcohols

The minimised alcohol geometries and the corresponding experimental values are shown in Table 6.3. The only gas phase electron diffraction data available for alcohols, to our knowledge, is that for methanol by Kubo and Kimura.[5] Microwave structure determinations for alcohols are both more common and more recent, and it is these that we have chosen to use for our comparisons.

When the ether **C–O** bond parameters were used, the calculated **C–O** bond lengths for the three alcohols (methanol, ethanol and *i*-propanol) were found to be too short by about 0.02 Å). This therefore required a further atom type (O_H) relating to a hydroxyl oxygen, which has a longer b_0 value for the **C–O** bond than the ether oxygen (**O**). (Why the **C–O** bond lengths for alcohols and ethers could not be fitted using the same parameters requires some explanation, since experimentally these bonds are found to be of very similar length. It is likely that the neglect of van der Waals interactions for hydroxyl hydrogens (see Chapter 5) means that the alcohol **C–O** bond is not stretched by the 1,4 interactions that occur for ether **C–O** bonds.)

The structure for methanol (with the exception of the **C–H** bonds; see Section 6.2) is in good agreement with experiment.

For ethanol, the derivation of the experimental geometry[25] deserves some comment. The values for the bond lengths **C–H**, **O–H** and angles **C–O–H** and **H–C–H**

Table 6.3 Comparison of Experimental and Calculated Alcohol Molecular Geometries[a]

Compound	Internal	Exptl.[b,c]	Calc.[c]	Diff.[d]	Reference
Methanol	**C–O**	1.425	1.426	0.001	r_0 (24)
	O–H	0.945	0.943	-0.002	
	C–H	*1.094*	*1.107*	*0.013*	
	C–O–H	108.5	108.5	0.0	
	H–C–H	*108.6*	*107.3*	*-1.3*	
Ethanol		*trans/gauche*	*trans/gauche*		r_0 (25)
	C–O	1.425/1.427	1.424/1.422	-0.001/-0.005	
	C–C	1.530	1.524	-0.006	
	C–C–O	107.3/112.4	110.3/110.6	3.0/-1.8	
	C–C–O–H	180/54±6	180/49	0/-5	
i-Propanol	**C–O**	1.434	1.418	-0.016	r_s (26)
	C–C	*1.523*	*1.527*	*0.004*	
	O–H	0.956	0.942	-0.014	
	C–H	*1.096*	*1.106*	*0.010*	
	C–O–H	107	106.3	-0.7	
	C–C–O	*108.7*	*108.9*	*0.1*	
	C–C–C	112.3	112.8	0.5	
	H–C–O–H	56	70	14	

a Bond lengths are in Å, Bond angles in degrees.

b Experimental values are derived from microwave spectral data.

c Values in *italics* indicate the internals which are assumed in the experimental model to be equivalent throughout the molecule (e.g. *all* **C–H** lengths in Methanol were assumed to be equal[24]). The appropriate calculated values are averaged to facilitate comparison.

d Diff = Calc - Exptl

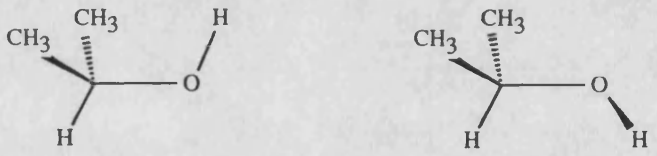
were assumed to be equal to those previously determined for methanol.[24] The values for the remainder of the geometry were then fitted to the rotational constants found by microwave. Both the *gauche* and *trans* forms were fitted, with the geometries of the two forms considered to differ only in the values of the **C–O** bond, the **C–C–O** angle and, of course, the **C–C–O–H** torsion angle.

Experimentally, the **C–O** bond is found to be slightly longer in the *gauche* form than the *trans*, although our calculations, like those of Rasmussen,[1] give the opposite result. This may be due to the attractive electrostatic interactions in both forcefields between the methyl carbon and the hydroxyl hydrogen . In addition, the neglect (in

our forcefield) of van der Waals terms for hydroxyl hydrogens will mean that the expected steric repulsion between these two atoms, which would counteract the electrostatic effects to some extent, will not be accounted for. (This problem is also discussed in Section 6.4.3, as it is also relevant to the energy difference between the two rotamers).

The **C–C–O** bond angle is also found to increase by 5.1° in going from the *trans* to the *gauche* form. The calculated values do not reproduce this, again, probably because the steric repulsions between the hydroxyl hydrogen and the methyl group are not accounted for.

Although not included in the parameter derivation, *i*-propanol is included here because it is the simplest secondary alcohol - a structural unit very common in carbohydrates. The values shown are those for the *gauche* conformation, which is experimentally found to be the more stable.[27]



trans gauche

i-Propanol

Assuming the experimental **C–O** bond length to be accurate, the calculated value is rather short. Otherwise, bond lengths and particularly the valence angles are well reproduced. The **H–C–O–H** torsion is slightly overestimated, although this may be due in part to the experimental uncertainty in determining hydrogen positions. The relative stabilities of the *gauche* and *trans* forms are discussed later in this chapter (Section 6.4.3).

6.2.4 Summary of Molecular Geometries

A summary of the differences between calculated and experimental structural data is made in Column A of Table 6.4. The figures given in the table are the average *absolute* deviations for all the bond lengths and valence angles given in Tables 6.1 to 6.3. The column headed 'No. of Internals' gives the number of each type of bond length or angle included in the sample.

Table 6.4 Summary of Geometrical Data : Average Absolute Deviations in Bond Lengths and Angles.

	A		B		
Internal	No. of Internals	Deviation[a]	No. of Internals	Deviation M&R[b]	Deviation This work[c]
Bond Lengths (Å)					
C–C	15(13)[d]	0.009 (0.004)[d]	7	0.003	0.003
C–H	15	0.009	8	0.013	0.009
C–O	9	0.005	4	0.002	0.002
O–H	2	0.008			
Angles (degrees)					
H–C–H	3	1.3	2	0.6	1.3
C–C–C	10	1.4	4	0.7	1.1
C–O–C	4	2.6	2	1.5	2.1
C–C–O	4	1.3	2	0.8	3.2
C–O–H	2	0.4	1	1.1	0.0

a Determined from Tables 6.1-6.3

b Deviations for the Melberg & Rasmussen forcefield were determined from calculated bond lengths and angles given in reference 1 and apply to the following compounds: Ethane, Propane, *i*-Butane, *n*-Butane *(trans)*, Neopentane, Cyclohexane, Dimethylether, 1,4-Dioxane, Methanol, Ethanol.

c This column shows the average deviations for the compounds listed in note *b* given by *our* forcefield.

d Excluding Tri-(*t*-butyl)-methane (which has highly strained **C–C** bonds).

The average deviation in **C–C** bond length is 0.009 Å, which seems rather large. If, however, the values for tri-(*t*-butyl)-methane are omitted (which gave an overestimated **C–C** bond length because of steric crowding) the average deviation falls to only 0.004 Å; a far more reasonable value. The **C–O** bonds are also reproduced to a similar level of accuracy, while the deviations in **C–H** and **O–H** bond lengths can be at least partly attributed to the large experimental error that arises in these values.

The valence angles are again reasonably well reproduced, although the large average deviation in the **C–O–C** angle is conspicuous, for reasons discussed earlier (Section 6.2.2).

Although an exhaustive comparison of our forcefield with others has not been carried out, it is instructive to look at how our geometries compare with those of the Melberg and Rasmussen forcefield[1] which was also developed specifically for modelling carbohydrates. In determining their forcefield, Melberg and Rasmussen (M & R) used many of the same model compounds as we have used. Column B of Table 6.4 compares the average deviations for those model compounds common to both studies. Bond length errors can be seen to be very similar for both forcefields, and although the M & R forcefield is generally better at reproducing valence angles, it too can be seen from the table to have the greatest deviations for the **C–O–C** valence angle.

It should be noted that the Melberg and Rasmussen forcefield was derived largely with the emphasis on fitting structure (rather than vibrational frequencies, rotational barriers and crystal properties as well; as was our intention). In this respect the M & R forcefield does remarkably well, especially when considering the simplicity of the potential energy function, and the small number of parameters used.

6.3 Vibrational Frequencies

It is important for a forcefield to be able to reproduce vibrational frequencies reasonably well if it is to be used to predict properties concerned with molecular motion. We consider vibrational frequencies to be one of the strong points of the forcefield described here, and its success is due in large measure to the use of cross-terms in the potential energy function.

The force constant parameters were optimised using the least-squares procedure (described in Chapter 4) to fit the vibrational spectra of seven molecules: ethane, propane, cyclohexane, dimethylether, 1,4-dioxane, methanol and ethanol. All the assigned fundamental frequencies of these molecules (with the exception of the **C–H** and **O–H** stretching frequencies) were included in the optimisation. The **C–H** and **O–H** stretching frequencies were not included because they are not important in determining conformational motions. The parameters for the **C–H** and **O–H** bonds (K_b, b_0 and α) were therefore transferred directly from the VFF[28] and were not adjusted further. The worst error in the **C–H** and **O–H** stretching region occurs in ethane (-79 cm^{-1}) and equates to an error of only 2.7% of the frequency value.

Five other molecules not used in the optimisation of the parameters are included here to demonstrate the transferability of the parameters. These molecules are: *n*-butane, *i*-butane, diethylether, 1,2-dimethoxyethane and ethylmethylether. The calculated vibrational frequencies for all 12 molecules studied, together with their symmetry species and corresponding experimental values are shown in Table 6.5 (1-12). The references from which the experimental values were obtained are indicated in the table.

Experimental errors for vibrational frequencies are difficult to quantify, but separate determinations carried out on the same molecule (see those cited for 1,4-dioxane) can often show frequency differences of up to 20 cm^{-1}, or even of 100 cm^{-1} where there is a difference of opinion over the assignment of a normal mode. Highly

flexible molecules present the greatest problems, as often the fundamental modes of the particular conformation of interest have to be identified from a complex spectrum containing absorptions from all the other conformations present. From these considerations, a low estimate for experimental error would therefore be 15 cm^{-1}.

Quite often, not all the fundamental frequencies are observed experimentally. This can occur for a number of reasons: weakly absorbing modes can be obscured by stronger modes of a similar frequency, or modes of a particular symmetry species may be inactive in IR spectroscopy, or Raman, or both.

We have chosen not to give specific assignments to the frequencies in Table 6.5, as in all but the most simple molecules the modes are not *pure* (in the sense that they correspond to the motion of a particular internal coordinate) but rather are a complex mixture of internal motions. In general terms, however, the deformations of the following internal coordinates can be considered to contribute to frequencies in the corresponding ranges:

C–C–H, H–C–H	1100-1500 cm^{-1}
C–C, C–O	500-1300 cm^{-1}
C–C–C, C–C–O, C–O–C	300-600 cm^{-1}
Torsions	0-500 cm^{-1}

However, it should be stressed that, even in medium-sized molecules (and especially ring systems) a high degree of coupling does occur resulting in impure modes.

Table 6.5 Comparison of Calculated and Experimental Vibrational Frequencies (cm^{-1})

1. Ethane (D_{3d})

Symm.	Calc.	Expt.	Devn.
E_g	2974	2969	5
E_u	2967	2985	-18
A_{1g}	2875	2954	-79
A_{2u}	2862	2896	-34
E_g	1461	1468	-7
E_u	1443	1469	-26
A_{1g}	1390	1388	2
A_{2u}	1375	1379	-4
E_g	1182	1190	-8
A_{1g}	995	995	0
E_u	834	822	12
A_{1u}	296	289	7

Experimental data from reference 29.

2. Propane (C_{2v})

Symm.	Calc.	Expt.	Devn.
A_1	2974	2977	-3
B_2	2973	2973	0
B_1	2971	2968	3
A_2	2971	2967	4
B_2	2956	2968	-12
A_1	2900	2962	-62
B_1	2871	2887	-16
A_1	2868	2887	-19
B_2	1467	1472	-5
B_1	1464	1464	0
A_2	1460	1451	9
A_1	1452	1476	-24
A_1	1431	1462	-31
B_1	1402	1378	24
A_1	1391	1392	-1
B_1	1348	1338	10
A_2	1303	1278	25
B_2	1176	1192	-16
A_1	1146	1158	-12
B_1	1036	1054	-18
B_1	959	922	37
A_2	943	940	3
A_1	872	869	3
B_2	746	748	-2
A_1	369	369	0
B_2	265	268	-3
A_2	226	216	10

Experimental data from reference 29.

Table 6.5 (continued)

3. *n*-Butane *(trans)* (C_{2h})

Symm.	Calc.	Expt.	Devn.
A_g	2973	2965	8
B_u	2973	2966	7
B_g	2973	2965	8
A_u	2972	2966	6
B_g	2959	2912	47
A_u	2954	2920	34
A_g	2904	2872	32
B_u	2894	2875	19
A_g	2869	2853	16
B_u	2869	2861	8
A_g	1469	1468	1
A_u	1466	1459	7
B_g	1462	1455	7
B_u	1454	1468	-14
B_u	1432	1451	-19
A_g	1429	1441	-12
A_g	1413	1377	36
B_u	1389	1378	11
A_g	1380	1360	20
B_g	1362	1303	59
B_u	1296	1291	5
A_u	1296	1258	38
B_g	1159	1181	-22
A_g	1140	1150	-10
A_g	1036	1058	-22
B_u	1035	1009	26
A_u	1005	948	57
B_u	989	964	25
A_g	857	837	20
B_g	815	805	10
A_u	716	732	-16
A_g	390	430	-40
B_u	288	*267*	21
B_g	261	*254*	7
A_u	228	*219*	9
A_u	136	*142*	-19

Experimental data from reference 30 (values in *italics* are from reference 31).

4. *i*-Butane (C_{3v})

Symm.	Calc.	Expt.	Devn.
A_1	2974	2965	9
E	2974	2958	16
E	2971	2951	20
A_2	2969	-	-
A_1	2919	2904	15
E	2872	2879	-7
A_1	2871	2879	-8
E	1484	1475	9
A_1	1481	1468	13
E	1458	1459	-1
A_2	1450	-	-
E	1421	1365	56
A_1	1408	1389	19
E	1345	1330	15
A_1	1186	1189	-3
E	1178	*1166*	12
A_2	1033	-	-
E	972	961	11
E	954	913	41
A_1	830	796	34
A_1	434	433	1
E	371	367	4
E	264	-	-
A_2	249	-	-

Experimental data from reference 32 (values in *italics* are from reference 33).

Table 6.5 (continued)

5. Cyclohexane (D_{3d})

Symm.	Calc.	Expt.	Devn.
A_{1g}	2970	2938	32
E_u	2965	2932	33
A_{2u}	2963	2934	29
E_g	2962	2926	36
A_{1g}	2912	2853	59
E_u	2907	2863	44
E_g	2900	2855	45
A_{2u}	2895	2855	40
A_{1g}	1449	1451	-2
E_u	1443	1454	-11
E_g	1439	1444	-5
A_{2u}	1436	1454	-18
A_{1u}	1386	-	-
E_u	1373	1350	23
A_{2g}	1369	-	-
E_g	1365	1348	17
E_g	1287	1267	20
E_u	1272	1259	13
A_{1u}	1163	-	-
A_{1g}	1150	1157	-7
A_{2g}	1132	-	-
A_{1u}	1052	-	-
E_g	1022	1029	-7
E_u	938	905	33
A_{2u}	873	-	-
E_u	858	862	-4
A_{1g}	811	802	9
E_g	772	785	-13
A_{2u}	532	524	8
E_g	473	425	48
A_{1g}	376	383	-7
E_u	247	248	-1

Experimental data from reference 34.

6. Dimethylether (C_{2v})

Symm.	Calc.	Expt.	Devn.
A_1	2977	2995	-18
B_1	2970	2995	-25
B_2	2968	2930	38
A_2	2966	-	-
B_1	2869	2820	49
A_1	2867	2820	47
B_2	1480	1462	18
B_1	1477	1464	13
A_2	1464	-	-
A_1	1457	1473	-16
A_1	1446	1452	-6
B_1	1434	1449	-15
A_1	1262	1248	14
B_2	1177	1180	-3
B_1	1174	1172	2
A_2	1154	1149	5
B_1	1073	1099	-26
A_1	910	926	-16
A_1	442	422	20
B_2	254	242	12
A_2	190	201	-11

Experimental data are average values taken from references 29, 35, 36 & 37.

Table 6.5 (continued)

7. Diethylether *(trans-trans)* (C_{2v})

Symm.	Calc.	Expt.	Devn.
A_1	2984	-	-
B_1	2983	-	-
B_2	2980	-	-
A_2	2980	-	-
B_2	2954	-	-
A_2	2952	-	-
A_1	2896	-	-
B_1	2893	-	-
A_1	2875	-	-
B_1	2875	-	-
B_1	1492	1484	8
A_1	1473	1490	-17
B_2	1464	1443	21
A_2	1463	1443	20
B_1	1450	1456	-6
A_1	1447	1456	-9
B_1	1399	1383	16
A_1	1388	1414	-26
B_1	1365	1351	14
A_1	1357	1372	-15
B_2	1352	1279	73
A_2	1320	-	-
A_1	1228	1168	60
B_2	1148	1168	-20
A_2	1142	1153	-11
B_1	1135	1120	15
B_1	1079	1077	2
A_1	1079	1043	36
B_1	943	935	8
A_1	903	846	57
B_2	828	822	6
A_2	811	794	17
A_1	424	440	-16
B_1	411	440	-29
B_2	278	245	33
A_2	264	231	33
A_1	196	208	-12
B_2	116	120	-4
A_2	99	120	-21

Experimental data from reference 29.

8. 1,4-Dioxane (C_{2h})

Symm.	Calc.	Expt.	Devn.
A_g	2978	2967	11
B_u	2974	2972	2
A_u	2973	2972	1
B_g	2970	-	-
A_g	2914	2855	59
A_u	2913	2865	48
B_g	2908	-	-
B_u	2904	2865	39
A_g	1471	1443	28
A_u	1467	1450	17
B_g	1466	1460	6
B_u	1459	1456	3
B_u	1389	1373	16
A_g	1373	1335	38
B_g	1367	1396	-29
A_u	1335	1368	-33
A_g	1309	1304	5
B_u	1274	1291	-17
A_u	1255	1257	-2
B_g	1238	1216	22
A_g	1153	1127	26
B_g	1149	1109	40
A_u	1120	1126	-6
A_u	1088	1085	3
A_g	1001	1015	-14
B_u	998	1050	-52
A_u	905	885	20
B_u	871	881	-10
A_g	867	835	32
B_g	836	853	-17
B_u	630	611	19
B_g	576	487	89
A_g	489	434	55
A_g	394	424	-30
B_u	276	275	1
A_u	266	286	-20

Experimental data are average values taken from references 35, 38, 39 & 40.

Table 6.5 (continued)

9. 1,2-Dimethoxyethane *(trans-trans-trans)* (C_{2h})

Symm.	Calc.	Expt.	Devn.
A_g	2973	-	-
B_u	2973	-	-
B_g	2967	-	-
A_u	2966	-	-
B_g	2964	-	-
A_u	2958	-	-
A_g	2907	-	-
B_u	2895	-	-
A_g	2867	-	-
B_u	2867	-	-
A_g	1497	1470	27
B_u	1476	1490	-14
B_g	1472	1450	22
A_u	1472	1451	21
A_g	1460	1470	-10
B_u	1460	1460	0
A_g	1442	1450	-8
B_u	1440	1459	-19
A_g	1401	1410	-9
B_g	1373	1270	103
A_u	1327	1286	41
B_u	1319	1338	-19
A_g	1257	1208	49
B_u	1226	1210	16
B_g	1165	1155	10
A_u	1164	1160	4
A_g	1103	1138	-35
B_g	1103	1092	11
B_u	1099	1122	-23
A_g	1079	1063	16
A_g	1026	996	30
B_u	954	938	16
A_u	808	823	-15
B_u	542	513	29
A_g	370	396	-26
A_g	328	-	-
A_u	257	-	-
B_g	220	-	-
B_u	165	-	-
A_u	146	-	-
B_g	136	-	-
A_u	77	-	-

Experimental data from reference 30.

10. Ethylmethylether *(trans)* (C_s)

Symm.	Calc.	Expt.	Devn.
A'	2983	-	-
A"	2980	-	-
A'	2974	-	-
A"	2967	-	-
A"	2953	-	-
A'	2894	-	-
A'	2875	-	-
A'	2868	-	-
A'	1487	-	-
A"	1472	-	-
A"	1463	-	-
A'	1462	-	-
A'	1449	-	-
A'	1441	-	-
A'	1390	1394	-4
A'	1362	1367	-5
A"	1338	1275	63
A'	1258	1219	39
A"	1165	1175	-10
A"	1144	1149	-5
A'	1117	1118	-1
A'	1096	1094	2
A'	1036	1019	17
A'	883	855	28
A"	819	820	-1
A'	454	472	-18
A'	294	308	-14
A"	276	238	38
A"	214	-	-
A"	113	-	-

Experimental data from reference 35.

Table 6.5 (continued)

11. Methanol (C_s)

Symm.	Calc.	Expt.	Devn.
A'	3632	3681	-49
A'	2968	3000	-32
A"	2967	2960	7
A'	2864	2844	20
A'	1460	1477	-17
A"	1443	1477	-34
A'	1438	1455	-17
A'	1364	1345	19
A"	1133	1165	-32
A'	1050	1060	-10
A'	1030	1033	-3
A"	303	295	8

Experimental data from reference 29.

12. Ethanol *(trans)* (C_s)

Symm.	Calc.	Expt.	Devn.
A'	3653	3676	-23
A'	2979	2989	-10
A"	2977	2989	-12
A"	2955	2949	6
A'	2895	2943	-48
A'	2873	2901	-28
A'	1478	1490	-12
A"	1460	1452	8
A'	1449	1452	-3
A'	1420	-	-
A'	1371	1394	-23
A'	1289	1241	48
A"	1269	-	-
A"	1143	1062	81
A'	1080	1089	-9
A'	1059	1033	26
A'	906	885	21
A"	817	801	16
A'	387	419	-32
A"	291	243	48
A"	267	201	66

Experimental data from reference 41.

6.3.1 Comparison of Calculated Frequencies with Experiment (Table 6.5)

Hydrocarbons

Ethane. For ethane, the A_{1g} mode at 995 cm^{-1} is almost a pure **C–C** stretch, and is reproduced exactly by the calculation. The lowest frequency (A_{1u}: calc. 296, expt. 289) is the **C–C** torsion and is reproduced well, as are all frequencies other than the **C–H** stretching modes (>2000 cm^{-1}).

Propane. There are two fairly pure **C–C** stretching modes in propane that arise from the two **C–C** bonds. These are symmetric (A_1: calc. 872, expt. 869) and antisymmetric (B_1: calc. 1036, expt. 1054). The single **C–C–C** valence angle gives a bending mode (A_1) at 369 cm^{-1} that the calculation reproduces exactly. The two lowest frequencies relate to the symmetric (A_2: calc. 226, expt. 216) and antisymmetric (B_2: calc. 265, expt. 268) torsional modes of the **C–C** bonds.

n-Butane. The vibrational spectra of this molecule has been resolved into the contributions from the *gauche* and *trans* forms. Only the more symmetric *trans* form is considered here. *Trans-n*-butane was not used in the least-squares optimisation of the parameters, but is nevertheless reasonably well fitted by the forcefield.

The **C–C** stretching modes are rather impure, but the **C–C–C** bending modes can be assigned as the (A_g: calc. 390, expt. 430) and (B_u: calc. 288, expt. 267) frequencies. There are three torsional modes in *n*-butane, all in good agreement with experiment. The central **C–C** bond is largely responsible for the lowest frequency mode (A_u: calc. 136, expt. 142) and the two terminal **C–C** bonds give rise to the next lowest modes (A_u: calc. 228, expt. 219) and (B_g: calc. 261, expt. 254)

i-Butane. In the C_{3V} point group, A_2 modes are not observed, being inactive in both IR and Raman spectroscopy. The A_1 mode at 796 cm^{-1} is slightly overestimated by

the calculation at 830 cm^{-1}. This mode is basically a symmetric stretch of the three **C–C** bonds. The lowest observed frequency (E: calc. 371, expt. 367) are degenerate **C–C–C** bending modes that are well reproduced.

Cyclohexane. The influence of the endocyclic valence angles of six-membered rings on the ring torsions means that the vibrations in cyclohexane are highly coupled and assignments therefore difficult to make. The frequency values (again, below 2000 cm^{-1}) are, however, generally very well fitted by the forcefield.

Ethers

Dimethylether. A number of vibrational studies of dimethylether have been made, all with very similar assignments and frequency values.[29,35-37] We have chosen an average of these frequency values for comparison.

The modes containing the largest **C–O** stretching contribution are the modes (A_1: calc. 910, expt. 926) and (B_1: calc. 1073, expt. 1099). The **C–O–C** bending vibration is the lowest A_1 mode (calc. 442, expt. 422). The two lowest modes are the symmetric and antisymmetric **C–O** torsional modes, (A_2: calc. 190, expt. 201) and (B_2: calc. 254, expt. 242).

Diethylether. The vibrational spectra of diethylether has been assigned for two out of the three possible conformers (*trans-trans* and *trans-gauche* forms).[29] The *trans-trans* form is the more symmetrical (C_{2V}) and lowest energy conformer.

The **C–O** and **C–C** stretching gives highly coupled vibrations, as do the **C–C–O** and **C–O–C** bending. Diethylether was not used in parameter optimisation.

1,4-Dioxane. This molecule was used in the optimisation of parameters and proved

one of the most difficult to fit. As for dimethylether, a number of studies have been made,[35, 38-40] and the values in Table 6.5(8) are an average of these frequency values. The average (absolute) error in frequencies (below 2000 cm^{-1}) is 23.2 cm^{-1}, which although reasonable, is the worst case of the ethers studied (see Table 6.6). The difficulties in fitting this molecule may be partly attributable to the *gauche effect*, which occurs in *gauche* **O–C–C–O** fragments (see Section 6.4.2). Two of these fragments are present in the dioxane ring.

1,2-Dimethoxyethane. Like 1,4-dioxane, 1,2-dimethoxyethane is a molecule that also exhibits the *gauche effect*. At room temperature, it exists in a mixture of conformations, with the *gauche* rotamer about the central bond being the most stable.[42] The vibrational spectra has been resolved into contributions from four conformational forms (*t-t-t, t-t-g, t-g-t* and *t-g-g*). The all-*trans* form has been used for comparison with calculation as it possesses the most symmetry. Even though not used in the optimisation process, the calculated frequencies are found to be in good agreement with experiment in all cases except one. This is the B_g vibration (calc. 1373, expt. 1270) which is found to be a mostly **C–C–H** bending mode. It may be that this discrepancy is an error of assignment rather than an error of the forcefield.

Ethylmethylether. Only one torsional mode for ethylmethylether has been observed, which according to the calculation is the **C–C** torsion (A": calc. 276, expt. 238). The **C–C–O** and **C–O–C** bending is highly coupled and give rise to the symmetric modes (A': calc. 294, expt. 308) and (A': calc. 454, expt. 472)

Alcohols

There are few well-assigned vibrational studies on alcohols, probably because the

bent **C–O–H** fragment means that the symmetry is restricted to a plane of reflection (C_s) at most. Only the vibrational frequencies of methanol and ethanol are considered here.

Methanol. The **C–O** torsional mode is reproduced well (A": calc. 303, expt. 295) as is the **C–O** stretching mode (A': calc. 1050, expt. 1060). **C–O–H** stretching, though mixed, contributes most strongly to the A' vibration at 1345 cm^{-1} (calc. 1364).

Ethanol. The vibrational spectra of this molecule has been measured in the vapour phase and in an argon matrix, and it is best interpreted as fitting the *trans* form.[41] For neither of the two torsional frequencies, **C–O** (A": calc. 267, expt. 201) and **C–C** (A": calc. 291, expt. 243), is the calculated value in good agreement with experiment by the standards set for the other molecules studied. Ethanol proved difficult to fit using the same parameters for the **C–C–O** valence angle as the ethers, but without more experimental data we were reluctant to introduce a specific set of **C–C–O** valence angle parameters for alcohols.

6.3.2 Summary of Vibrational Frequencies

Table 6.6 shows a summary of the deviations in vibrational frequencies for all the molecules studied. Only deviations in the skeletal vibrations (below 2000 cm^{-1}) are considered, and the overall deviations are expressed as an average of the *absolute* deviation values shown in Table 6.5 (1-12). Table 6.6 is divided into three sections, each dealing with one of the three classes of compounds studied; hydrocarbons, ethers and alcohols.

Generally, and perhaps not surprisingly, the simplest molecules in each class of compound are seen from Table 6.6 to give the best fit to experiment. This occurs for two reasons. Being small molecules, the vibrational modes are purer and therefore

Table 6.6 Summary of Vibrational Data : Average Absolute Deviations of Calculated Vibrational Frequencies (below 2000 cm^{-1})[a]

Compound	Point Group	Average[b] Dev. (cm^{-1})	Maximum Dev. (cm^{-1})	No. Freq.s
Hydrocarbons				
Ethane	D_{3d}	8.3	26	8
Propane	C_{2v}	12.3	37	19
n-Butane	C_{2h}	20.5	59	26
i-Butane	C_{3v}	16.9	56	13
Cyclohexane	D_{3d}	13.7	48	18
Total (hydrocarbons)		15.5		84
Ethers				
Dimethylether	C_{2v}	12.6	26	14
Diethylether	C_{2v}	21.6	73	28
1,4-Dioxane	C_{2h}	23.2	89	28
Ethylmethylether	C_s	17.5	63	14
1,2-Dimethoxyethane	C_{2h}	22.9	103	25
Total (ethers)		20.6		109
Alcohols				
Methanol	C_s	17.5	34	8
Ethanol	C_s	30.0	81	13
Total (alcohols)		25.2		21
Total (all molecules)		19.0		214

a These frequencies involve skeletal vibrations only (i.e. all except **C–H** stretching modes).

b Average deviations are calculated from the deviations given in Table 6.5 (1-12).

easier to optimise with the force constant parameters. Additionally, because the simpler molecules give rise to less complex spectra, the experimental frequency assignments will be more reliable.

Of the three classes of compound studied, the hydrocarbons give the best agreement with experiment, having an average deviation of only 15.5 cm^{-1} over 84 frequency values. This is likely to be due to the absence of lone-pairs of electrons (found in ethers and alcohols) which give rise to electronic effects that are difficult to account for in valence forcefields of this type.

Ethers formed the largest class studied, with 109 frequencies in total having an average deviation of 20.6 cm^{-1}.

Because only two alcohols were included, the relatively poor fit of ethanol, with an average deviation of 30.0 cm^{-1} gives a large overall deviation for the 21 alcohol frequencies.

Overall, the total number of frequencies used in the comparison was 214, which were reproduced with an average error of 19.0 cm^{-1}. This represents excellent agreement when compared to experimental error for vibrational frequencies, which (as discussed in Section 6.3) may be in the region of 15 cm^{-1}, or even larger in poorly assigned spectra.

Although the Melberg and Rasmussen forcefield[1] was not fitted to reproduce vibrational frequencies with precision, an indication of the performance of the M & R forcefield is given in Table 6.7. Three molecules were chosen, *n*-butane, 1,4-dioxane and ethanol, that gave the worst results for each class of compound using our forcefield. As can be seen from the table, the M & R forcefield is substantially poorer at reproducing vibrational spectra. This is undoubtably due to the absence of cross-terms from this forcefield, which are recognised to have a large effect on vibrational frequencies.[43]

Table 6.7 Deviations of Vibrational Frequencies[a] given by the Melberg and Rasmussen Forcefield.[1]

Compound	Point Group	Average Dev. (cm^{-1})	Maximum Dev. (cm^{-1})	No. Freq.s
***n*-Butane**	C_{2h}			
M & R		45.6	206	22
This Work		21.7	59	
1,4-Dioxane	C_{2h}			
M & R		56.6	167	28
This Work		23.2	89	
Ethanol	C_s			
M & R		51.5	106	13
This Work		30.0	81	

a Frequencies below 2000 cm^{-1} only (i.e. all except **C–H** stretching modes).

6.4 Rotameric Energies

This section is concerned with the ability of the forcefield to reproduce rotational barriers and rotameric energy differences of individual bonds. Preferences for a particular rotation about individual bonds have a large influence on the overall conformation of large flexible molecules.

Although parameter values were not generally optimised to fit rotameric energies; occasionally, where a calculated energy difference was too far in error, a value of the relevant V_1 torsion parameter had to be selected to reproduce the experimental value. In the majority of cases, however, this was not necessary and the V_1 values were left at their initial value (zero).

Experimental errors for rotameric energies vary widely because of the range of techniques used to determine them. The choice of experimental data used here was often limited by availability. It therefore comes from many different sources, including

vibrational spectroscopy, microwave, NMR and calorimetric data, and refers to compounds in both the vapour phase and condensed phases.

The calculated rotational barriers were calculated by the method described in Chapter 3.

The rotameric energies of 14 molecules in total were studied, comprising of 4 hydrocarbons, 6 ethers and 4 alcohols. Graphs showing how the calculated potential energy of each molecule varies with rotation about the bond in question are included in Appendix II. These graphs also show a breakdown of the total potential energy in terms of the components of the potential energy function (i.e. van der Waals energy, electrostatic energy, and so on).

6.4.1 Hydrocarbons

Table 6.8 shows how the calculated rotameric energies of hydrocarbons compare with experimental values.

Ethane. Ethane only has one conformation as all rotameric forms are equivalent. The rotational energy barrier of the **C–C** bond (2.9 kcal/mol) has been estimated from an IR study in the gas phase.[44] The calculated value can be seen from Table 6.8 to be in good agreement with experiment, with an error of 0.26 kcal/mol (9%).

From the plot of one of the **H–C–C–H** torsion angles (ϕ) versus energy (Fig. 1, Appendix II) the three-fold sinusoidal shape of the total energy can be seen, as might be expected from symmetry considerations. The rotational barrier is composed almost entirely from the three-fold torsional term (V_3) of the potential: the V_3 parameter, it should be noted, was not fitted to the rotational barrier but to the torsional frequencies of the hydrocarbons.

n-Butane. Two conformations exist for *n*-butane, the *gauche* and *trans* forms. Although the *trans* form is known to be the more stable form, the precise energy

Table 6.8 Comparison of Calculated and Experimental Rotational Energies of Hydrocarbons.

Compound	Torsion	Relative Energy[a] Calc.	Expmt.	Source
Ethane	**H-C-C-H**			
	staggered	0	0	
	eclipsed	2.64	2.9	IR Spect.[c]
n-Butane	**C-C-C-C**			
	trans (t)	0	0	
	gauche(g)	1.00	0.89 (0.6)	Raman Spect.[d] (*Ab Initio*)[e]
	$\Delta E^{\ddagger}(t \rightarrow g)$[b]	3.97	3.63	Raman Spect.[d]
	$\Delta E^{\ddagger}(g^{+} \rightarrow g^{-})$	6.15	4.52 (6.34)	Raman Spect.[d] (*Ab Initio*)[e]
2-Methylbutane	**$H\text{-}C_2\text{-}C_3\text{-}C_4$**			
	gauche (g)	0	0	
	trans (t)	0.63	0.81	Raman Spect.[f]
2,2-Dimethylbutane	**C-C-C-C**			
	staggered	0	0	
	eclipsed	7.20	5.2	^{1}H NMR[g]

a Energies are in kcal mol^{-1}
b $\Delta E^{\ddagger}$ represents the *barrier height* of the indicated transition relative to the lowest energy rotamer.
c From Reference 44. *d* From Reference 45. *e* From Reference 14. *f* From Reference 46.
g From Reference 47.

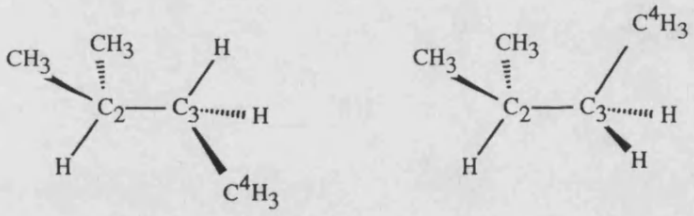
difference between the two forms is the subject of debate, with values ranging from 0.5 to 0.97 kcal/mol.[45] The experimental value shown in the table is one of the more recent measurements by Compton *et al.* From the calculation, we obtain a $\Delta E(g - t)$ value of 1.00 kcal/mol, which is in reasonable accord with this.

The graph of the torsion angle versus energy (Fig. 2, Appendix II) shows this energy difference to be accountable mostly to electrostatic energy. This is caused by

the electrostatic repulsion between the two terminal methyl groups, which are closer together in the *gauche* form.

The *trans-gauche* energy barrier also agrees with experiment, although the *gauche-gauche* barrier seems rather too large. (It is, however, in in close agreement with high level basis set *ab initio* calculations[14]). The graph indicates that there are three main contributions to the *gauche-gauche* energy barrier; the torsional term, the electrostatic term (for the reasons given above) and angle strain, caused by the **C–C–C** angles opening as the terminal methyls are eclipsed. The fact that the **C–C–C** valence angles open demonstrates the difference between rigid geometry and flexible geometry calculations like those reported here. Rigid geometry calculations generally predict bigger rotational barriers than flexible geometry, because the valence angles cannot open to relieve 1,4 steric clashes.[48] Similarly to our calculations, fully optimised *ab initio* calculations also indicate the **C–C–C** angles to open in the fully eclipsed form of *n*-butane.[14]

2-Methylbutane. For this molecule, the torsion has been defined by the **H–C_2–C_3–C_4** dihedral angle.



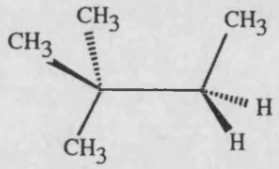
gauche trans

2-Methylbutane

The calculated rotameric energy difference is in good agreement with experiment, favouring the *gauche* form energetically. (This is because the *'trans'* form has two *gauche* methyl-methyl interactions as opposed to one in the *'gauche'* form).

Interestingly, Figure 3 (Appendix II) shows the van der Waals energy to be *stabilising* with respect to the *trans* form, but this is outweighed by a larger destabilising electrostatic contribution.

2,2-Dimethylbutane. As all rotamers are equivalent for this molecule, only the rotational barrier is of interest.



2,2-Dimethylbutane

Table 6.8 shows the calculated barrier to be overestimated when compared with experiment, which was determined by ^{1}H dynamic NMR spectroscopy.[47]

The energy barrier is caused by the eclipsing of terminal methyl groups, and is therefore analogous to the *gauche-gauche* barrier in *n*-butane, which was also overestimated. This may therefore indicate a weakness of the forcefield in this area.

6.4.2 Ethers

Table 6.9 shows how calculated rotameric energies of ethers compare with experimental values.

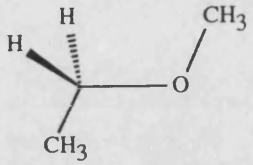
Dimethylether. The rotational energy barrier is slightly underestimated by the calculation. The plot of torsion angle versus energy (Fig. 5, Appendix II) shows the energy barrier to have two main components; the V_3 torsional term (fitted to the torsional frequency modes of ethers) and angle strain. The latter of these is caused by the opening of the **C–O–C** angle as the **H–C–O–C** torsion becomes eclipsed.

Table 6.9 Comparison of Calculated and Experimental Rotameric Energies of Ethers.

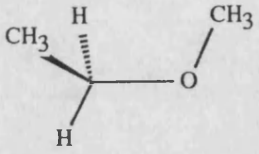
Compound	Torsion	Relative Energy[a] Calc.	Expmt.	Source
Dimethylether	**H-C-O-C**			
	staggered	0	0	
	eclipsed	2.11	2.7	Microwave[c]
Ethylmethylether	**C-O-C-C**			
	trans (t)	0	0	
	gauche (g)	1.41	1.11-1.5	Vibrational Spect.[d,e]
	$\Delta E^{\ddagger}(t \rightarrow g)$[b]	2.24	2.93	Vibrational Spect.[d]
	$\Delta E^{\ddagger}(g^{+} \rightarrow g^{-})$	6.89	4.07	Vibrational Spect.[d]
Diethylether	**C-O-C-C**			
	trans-trans	0	0	
	gauche-trans	1.31	1.4	IR Spect.[f]
i-Propylmethylether	**H-C-O-C_{Me}**			
	gauche (g)	0	0	
	trans (t)	2.34	2.2 (2.4)	^{13}C NMR[g](Raman Spect.)[h]
	$\Delta E^{\ddagger}(g^{+} \rightarrow g^{-})$	0.68	1.2	Semiempirical estimate[g]
	$\Delta E^{\ddagger}(g \rightarrow t)$	5.80	5.8	Semiempirical estimate[g]
t-Butylmethylether	**C-C-O-C**			
	staggered	0	0	
	eclipsed	3.31	3.57	IR Spect. (gas)[i]
	H-C-C-O			
	staggered	0	0	
	eclipsed	5.24	4.71	IR Spect. (solid)[i]
1,2-Dimethoxyethane	**O-C-C-O**			
	gauche (g)	0	0	
	trans (t)	0.36	0.5	^{13}CH NMR (nonpolar solvent)[j]

a Energies are in kcal/mol.
b $\Delta E^{\ddagger}$ represents the *barrier height* of the indicated transition relative to the lowest energy rotamer.
c From reference 22. *d* From reference 49. *e* From reference 50. *f* From reference 51. *g* From reference 52. *h* From reference 53. *i* From reference 54. *j* From reference 42.

Ethylmethylether. This molecule can be considered as analogous to *n*-butane, but with a central **C–O** bond rather than a **C–C**.



trans



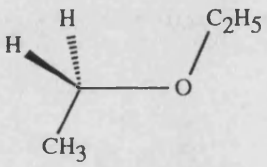
gauche

Ethylmethylether

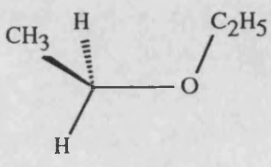
The *gauche-trans* energy difference is very well matched by calculation, falling within the range of the experimental values. The energy difference is larger than for *n*-butane, as the shorter **C–O** bond causes the terminal methyl groups to be brought even closer, so giving greater repulsive forces. This also results in a greater dihedral angle at the *gauche* minimum for ethylmethylether than for *n*-butane (see Sections 6.2.1 and 6.2.2).

Similarly to *n*-butane, the *trans-gauche* barrier for ethylmethylether is in good agreement with experiment, but the *gauche-gauche* barrier is again overestimated.

Diethylether. The **C–O** bonds in diethylether behave almost identically to the central **C–O** bond in ethylmethylether, as can be seen by comparing the rotational barrier plots (Fig. 6 & 7, Appendix II). The *gauche-trans* energy difference value is very similar to that for ethylmethylether both experimentally and in the calculation.



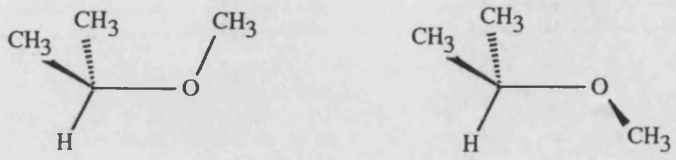
trans



gauche

Diethylether

i-Propylmethylether As ethylmethylether is analogous to *n*-butane, so *i*-propylmethylether is to 2-methylbutane. Like 2-Methylbutane, and for the same reasons, the *gauche* form is found to be the more stable.



trans gauche

i-Propylmethylether

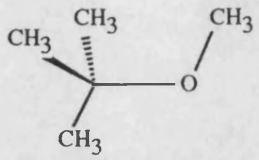
The two experimental values for this energy difference are in very close agreement: one was determined by NMR in cyclohexane solution,[49] the other by Raman spectroscopy on molecules trapped in an argon matrix.[50] The calculated energy difference also matches these values well.

Semiempirical estimates by Durig and Compton[49] (made on the basis of their NMR experiments) are shown in Table 6.9 for the two rotational barriers. These values are both in good agreement with our calculated values.

From the rotational barrier plot (Fig. 8, Appendix II) the largest component to the *gauche-trans* energy difference (and the *gauche-trans* barrier) is the angle term. As for ethylmethylether, this is because the **C–O–C** and **O–C–C** angles open up to relieve the repulsive methyl-methyl interactions.

t-Butylmethylether. Two rotational barriers have been determined for this molecule

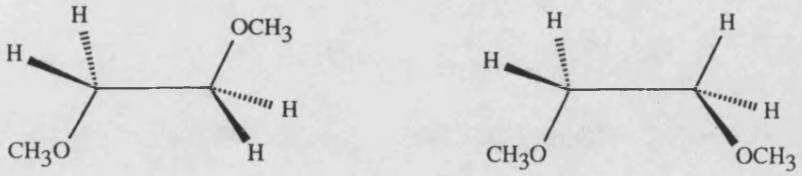
corresponding to the **C–C** torsions of the *t*-butyl group, and that of the **C–O** bond.



t-Butylmethylether

Both calculated barriers are in reasonable agreement with experimental values.

1,2-Dimethoxyethane. This molecule is unusual in that the *gauche* rotamer is found to be favoured by experiment.[42] This is contrary to what would be expected from electrostatic and steric considerations, which would indicate the oxygens to repel each other strongly, thus favouring the *trans*.



trans gauche

1,2-Dimethoxyethane

In fact, the *gauche* form is favoured because of an unusual stereoelectronic effect known as the *gauche effect* that occurs in compounds containing vicinally disubstituted electronegative atoms (**O**, **F**, **N** etc.).[23,55,56]

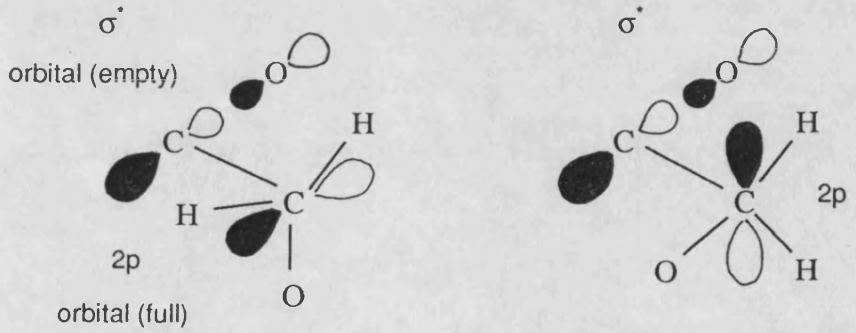
Initial calculations for 1,2-dimethoxyethane proved to be in error, showing, as expected, the *trans* form to be more stable by 4.6 kcal/mol, because of the

electrostatic repulsions between the two oxygens. The mechanism of the *gauche effect* is still under investigation, and so a proper molecular mechanics formalism for it must wait for the time being. However, a fairly crude (but pragmatic) solution to the problem has been made here, by using a large negative V_1 term for the **O–C–C–O** torsion (V_1 = -3.55 kcal/mol). A negative V_1 term gives a maximum at 180° and a minimum at 0°, and thus destabilises the *trans* form relative to the *gauche*. By adjusting the magnitude of the V_1 term, the experimental preference for the *gauche* from can be reproduced. (This V_1 term was also included in the calculations for 1,4-dioxane and 1,2-dimethoxyethane presented earlier in this chapter.)

A plot of the variation of the total energy with the **O–C–C–O** torsion angle is shown in Figure 11 (Appendix II). The large electrostatic term favouring the *trans* rotamer is clearly evident. The torsion term (including *both* the V_1 and V_3 terms) is responsible for raising the energy of the *trans* form slightly above that of the *gauche*.

The preference for the *gauche* rotamer was found by NMR to increase with solvent polarity.[42] In Table 6.9, we have chosen the ΔE value obtained in the least polar solvent (C_6D_{12}) as it relates more closely to our calculations on the isolated molecule.

[Note: A mechanism that has been put forward to explain the gauche effect involves a stabilising conjugation between bonding electrons in each methylene fragment with the σ orbital of the **C–O** bonds*[55]



$\phi_{OCCO} = 90°$ $\phi_{OCCO} = 180°$

This conjugation, and hence the stabilisation, will be at a maximum when ϕ_{OCCO} = 90° (pseudo-gauche) and falls to zero at ϕ_{OCCO} = 180° (trans). This mechanism is exactly analogous to that of the anomeric effect discussed in the next chapter. If this mechanism can be verified by experimental data (observed changes in bond length, and so forth) the torsion cross-term used to reproduce the anomeric effect could also be used to account for the gauche effect.]

6.4.3 Alcohols

Table 6.10 shows how calculated rotameric energies of alcohols compare with experimental values.

Methanol. The rotational barrier of methanol is fitted almost exactly by the calculation. The rotational barrier plot (Fig. 12, Appendix II) shows that the barrier is composed almost totally from the torsional (V_3) term. (The value of V_3 was fitted to the torsional frequency of methanol).

Ethanol. The experimental data on the *gauche-trans* energy difference is somewhat inconclusive. Although Perchard and Josien[61] concluded from their vibrational analysis of gaseous ethanol that only the *gauche* rotamer was present, microwave studies[25,57,62] and another vibrational study[41] indicated the *trans* form to be more stable. This also concurs with *ab initio* calculations.[63,64] The range of values of the energy difference $\Delta E(g\text{-}t)$ varies in these studies from 0.12-0.82 kcal/mol. We have chosen to use the lowest of these values, as it relates to the most recent determination by Kakar and Quade.[57]

As can be seen from Table 6.10, contrary to experiment, the *gauche* form is found to be the more stable rotamer. From Figure 13a (Appendix II) it is evident that this is due to a large destabilisation of the *trans* rotamer due to the electrostatic energy. This is a result of the electrostatic repulsions between the hydroxyl hydrogen

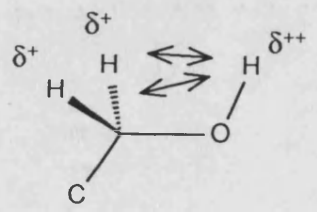
Table 6.10 Comparison of Calculated and Experimental Rotameric Energies of Alcohols.

Compound	Torsion	Relative Energy[a] Calc.	Expmt.	Source
Methanol	**H-C-O-H**			
	staggered	0	0	
	eclipsed	1.04	1.07	Microwave[c]
Ethanol	**C-C-O-H**			
	gauche(g)	0	0	
	trans (t)	1.36 (-0.23)[h]	-0.12	Microwave[d]
	$\Delta E^{\ddagger}(g^{+}\rightarrow g^{-})^{b}$	0.67 (1.31)[h]	1.2	Microwave[e]
i-Propanol	**H-C-O-H**			
	trans (t)	0	0	
	gauche (g)	1.18 (-0.15)[h]	-0.10	IR/Raman[f]
t-Butanol	**C-C-O-H**			
	staggered	0	0	
	eclipsed	1.67 (1.67)[h]	0.9	Calorimetric data[g]

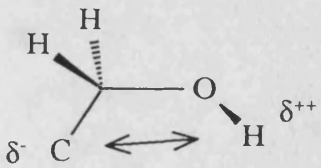
a Energies are in kcal mol^{-1}
b $\Delta E^{\ddagger}$ represents the *barrier height* of the indicated transition relative to the *gauche* rotamer.
c From reference 24. *d* From reference 57. *e* From reference 58. *f* From reference 59. *g* From reference 60.

h Values in parentheses are for calculations for which $V_{1\,CCOH}$ = 1.00 kcal/mol

and the methylene hydrogens in the *trans* form, and an electrostatic *attraction* between the hydroxyl hydrogen and the methyl carbon in the *gauche* form.



trans

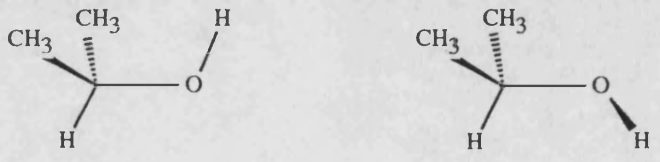


gauche

The Melberg and Rasmussen forcefield also predicts the *gauche* form to be favoured, presumably for similar reasons.[1] Because van der Waals forces are not considered for hydroxyl hydrogens (see Chapter 5) the expected steric repulsion between the methyl group and the hydroxyl hydrogen in the *gauche* form is not accounted for by our forcefield. Even so, it is unlikely if a van der Waals term could account for the necessary energy to stabilise the *gauche* form relative to the *trans* (>1.2 kcal/mol). Repeating the calculations using the van der Waals parameters (r^*_{ii} and ε_{ii}) of an aliphatic hydrogen (**H**) for those of the hydroxyl hydrogen ($\mathbf{H_O}$) still gave the *gauche* form to be the more stable (by 0.4 kcal/mol). The hydroxyl hydrogen would therefore have to be substantially 'larger' than an aliphatic hydrogen in order to reproduce experiment, which is contrary both to chemical intuition and the crystal simulations of Lifson *et al.*,[65] discussed in Chapter 5.

As for 1,2-dimethoxyethane, the problem can be solved by recourse to the V_1 torsional term. Using a V_1 value of 1.00 kcal/mol for the **C–C–O–H** torsion can be seen in Figure 13b (Appendix II) to raise the energy of the *gauche* form above that of the *trans* by 0.23 kcal/mol, in reasonable agreement with experiment. The total energy (in Figure 13b) now resembles a simple three-fold barrier, with the *trans-gauche* and *gauche-gauche* barriers having very similar values (1.47 kcal/mol and 1.31 kcal/mol respectively). The experimental value of the *gauche-gauche* barrier has been determined[58] and is now in good agreement with our calculated value.

i-Propanol. For *i*-propanol, the *gauche* and *trans* rotamers are defined relative to the **H–C–O–H** torsion angle.



trans gauche

i-Propanol

Vibrational spectroscopy of a dilute solution in CCl_4 indicates that the *gauche* form is slightly more stable by 0.10 kcal/mol.[59] Again, calculated values were initially in error, indicating the *trans* form to be the more stable rotamer. By adopting the same V_1 parameter that was used in ethanol, this error is rectified, resulting in a more stable *gauche* form. This is further support, therefore, for the use of the V_1 parameter for the **C–C–O–H** torsion.

The plots of the torsion angle versus energy are shown in Figures 14a & 14b (Appendix II). These relate to the calculations performed with and without the **C–C–O–H** V_1 (= 1.00 kcal/mol) parameter respectively.

t-Butanol. A value of 0.9 kcal/mol for the rotational barrier in *t*-butanol has been estimated from vapour heat capacity measurements and calorimetric entropy data.[60] This is may be a rather low estimate, when compared to the experimental *gauche-gauche* energy barrier in ethanol of 1.2 kcal/mol.[58] This barrier is also caused by an eclipsing of a methyl group with a hydroxyl hydrogen, and would therefore be expected to be of a similar value. In this light, the calculated barrier height of 1.67 kcal/mol does not seem too unreasonable.

6.5 Conformational and Configurational Energies

The ability to predict, with reasonable confidence, the relative energies of different conformations, is essential for a forcefield that is to be used in conformational analysis. Relative energies of conformations that differ only in the rotation about a bond (rotamers) were discussed in the previous section. Here, we consider the conformational (and configurational) energies of ring systems, which, for obvious reasons will be important in the extension of the forcefield to carbohydrates.

Configurations are structural isomers that differ not in connectivity but in the *spatial* arrangements of atoms. For example, the α- and β- forms of glucose are

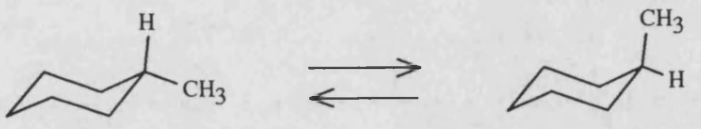
different *configurations*, as are the L and D forms of an amino acid. Our use of configurational energy differences alongside conformational ones deserves special comment. Although two configurations cannot interconvert in the way conformers do, the topological connectivity (or *constitution* as it is sometimes called[3]) is the same for both. Thus, the two configurations will have the same number and types of chemical bonds, and the energy difference will therefore be the result of a difference in intramolecular strain energy. This is, of course, what the forcefield seeks to represent.

Conformational and configurational energies are obtained from a number of experimental techniques. Where ΔH values have been determined, these values are directly comparable with energy differences found by calculation. In some cases, however, only ΔG values were available, and it should be born in mind that these contain an entropic component that is not included in the calculated values.

Table 6.11 shows a summary of how the calculated energy differences ($\Delta E_{conf.}$) compare with the corresponding experimental values.

6.5.1 Hydrocarbons

Methylcyclohexane. For methylcyclohexane, the axial-equatorial energy difference (ΔH) has been determined by variable temperature NMR in $CFCl_3/CDCl_3$ solution.[66] The calculated ΔE value slightly overestimates the preference for the equatorial form.



equatorial axial

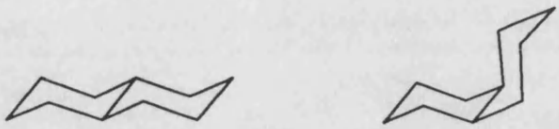
Methylcyclohexane

Table 6.11 Comparison of Experimental and Calculated Conformational and Configurational Energy Differences

Compound	$\Delta E_{conf.}$ (kcal/mol) Calc.	Expt.	Source
Hydrocarbons			
Methylcyclohexane			
ΔE(*axial - equatorial*)	2.17	1.75	^{13}C NMR[a]
Decalin			
ΔE(*cis - trans*)	2.76	2.69	Difference in ΔH_c values[b,c]
1,4-Dimethylcyclohexane			
ΔE(*cis - trans*)	2.19	1.89	Difference in ΔH_c values[b,c]
Cyclohexane			
ΔE(*twist-boat - chair*)	8.63	5.5	High-vacuum deposition/IR[d]
Ethers			
Methoxycyclohexane			
ΔE(*axial - equatorial*)	0.72	0.71	^1H NMR[e]
2,2-Dimethylmethoxycyclohexane			
ΔE(*axial - equatorial*)	1.32	0.54	^1H NMR (ΔG value)[f]
Tetrahydrofuran			
$\Delta E(C_s - C_2)$	0.81	0.16	Microwave[g]
$\Delta E(C_{2v} - C_2)$	2.87	3.46	Microwave[g]
trans-2-Methoxy-*cis*-decalin			
ΔE(*axial - equatorial*)	0.46	0.20	^1H NMR (ΔG value)[h]
cis-2-Methoxy-*cis*-decalin			
ΔE(*axial - equatorial*)	1.00	1.3	^1H NMR (ΔG value)[i]

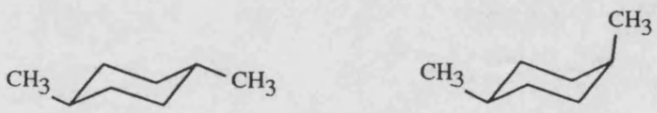
a From reference 66. *b* From reference 67. *c* From reference 68. *d* From reference 69. *e* From reference 70. *f* From reference 71. *g* From reference 72. *h* From reference 73. *i* From reference 74.

Decalin. The enthalpy difference between the *cis* and *trans* configurations of decalin can be determined from the difference in their heats of combustion.[67,68] This energy difference (favouring the *trans* form) is very well reproduced by the forcefield.



trans-decalin cis-decalin

1,4-Dimethylcyclohexane. Heats of combustion values were again used to determine the energy difference between *cis* and *trans*-1,4-dimethylcyclohexane. These differ by the orientation of one methyl group; axial in the *cis* and equatorial in the *trans* form:



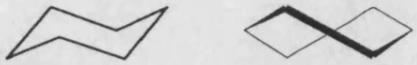
trans- cis-

1,4-Dimethylcyclohexane

As for methylcyclohexane, the calculated energy difference is again slightly overestimated.

Cyclohexane. There are two energy minima for cyclohexane, found both experimentally and by our calculations. The much higher energy twist-boat form has been

identified using IR by trapping cyclohexane vapour (at 1073 K) on a CsI plate (cooled to 20 K).[69]



chair twist-boat

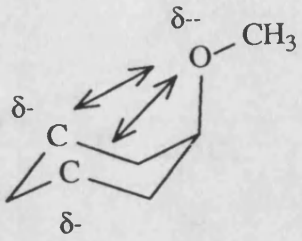
Cyclohexane

The rate at which the twist-boat decayed to the chair form was then used to estimate the energy difference between the two conformers. The discrepancy between the calculated and observed values of some 3 kcal/mol may be due in part to the unconventional experimental method used. However, the calculated energy for the twist-boat form may also be overestimated due to the near-eclipsed **C–C–C–C** torsion angles. Note that the forcefield also gave an exaggerated *gauche-gauche* barrier for *n*-butane, which also requires the eclipsing of a **C–C–C–C** torsion (see Section 6.4.1).

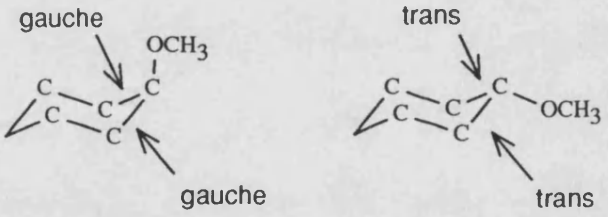
6.5.2 Ethers

Methoxycyclohexane. Experiment has shown that the axial form of methoxycyclohexane is 0.71 kcal/mol higher in energy than the equatorial conformer.[70] Initial calculations on methoxycyclohexane gave an axial-equatorial energy difference that was much too large (by >2 kcal/mol) when compared with the experimental value. This was found to be caused by large repulsive electrostatic interactions between the ring

carbons and the oxygen atom occurring in the axial form.

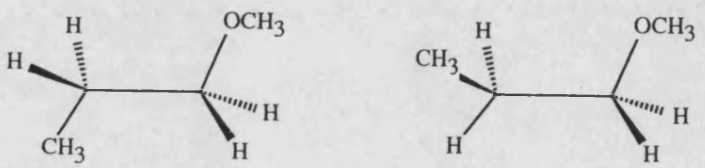


As it seemed unlikely these electrostatic interactions were so far in error, this led us to examine the *gauche/trans* relationship of the **C–C–C–O** fragment. In simple terms, axial and equatorial methoxycyclohexane can be considered to differ only in the rotations of two such **C–C–C–O** fragments. Both are *gauche* in the axial form and *trans* in the equatorial form.



axial equatorial

In order to study this fragment, n-propylmethylether was chosen as a simple model compound.



trans gauche

n-Propylmethylether

Although no experimental data for the *gauche-trans* energy difference for this molecule could be found, it is commonly assumed that the **C–C–C–O** fragment favours the *trans* form by about 0.4 kcal/mol.[75] This value was very much lower than our initial calculated value of 1.45 kcal/mol. High level basis set *ab initio* results were at even greater odds with our calculated value, indicating *n*-propylmethylether to actually favour the *gauche* form.[76] An analogous fluoro- compound, 1-fluoropropane, contains an electronically similar fragment **C–C–C–F**, which was also found by *ab initio* to favour the *gauche* form[23] for stereoelectronic reasons similar to those proposed for the *gauche effect* (described in Section 6.4.2).

In our opinion, it seems likely that a stereoelectronic effect, similar to the *gauche effect* although not as pronounced, also exists in the **C–C–C–O** fragment, causing a degree of stabilisation in the *gauche* conformer. Since this is not accounted for in our forcefield, the electrostatic effects were found to dominate, causing the *gauche-trans* energy difference to be overestimated.

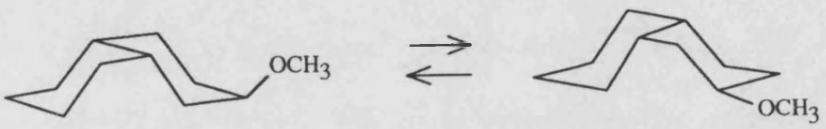
As in the cases of 1,2-dimethoxyethane and ethanol, a solution that gave the correct conformational energy difference was to use a non-zero V_1 parameter for the **C–C–C–O** torsion. By selecting a value for V_1 for this torsion of -0.88 kcal/mol, the observed axial-equatorial energy difference in methoxycyclohexane could be reproduced almost exactly (see Table 6.11).

The *gauche-trans* energy difference for *n*-propylmethylether was also recalculated, giving a revised energy difference of only 0.18 kcal/mol in favour of the *trans* form.

2,2-Dimethylmethoxycyclohexane. A free energy difference between the axial and equatorial forms has been determined by NMR in a dilute solution of CS_2.[71] The calculated energy difference is slightly too large (by 0.78 kcal/mol) but this may be due to entropic factors.

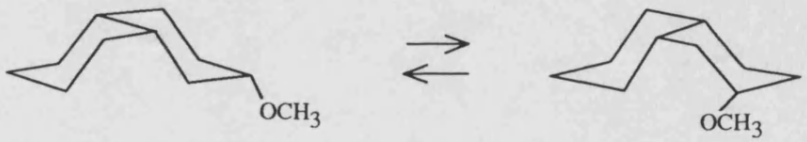
Tetrahydrofuran. Estimates from microwave studies indicate the barrier to pseudorotation (which requires a C_s transition state) to be 57 cm^{-1} (0.16 kcal/mol).[72] The barrier to planarity (i.e. having the C_{2v} point group) is also estimated at 1220 cm^{-1} (3.46 kcal/mol). Calculated values are in reasonable accord with these estimates.

trans- and cis-Methoxy-cis-decalin. The flexibility of *cis*-decalins means that these two compounds each have two well-defined conformations. The relative free energy differences of these conformations have been studied by NMR.[73, 74]



I (OCH_3 axial) II (OCH_3 equatorial)

trans-2-Methoxy-*cis*-decalin



III (OCH_3 equatorial) IV (OCH_3 axial)

cis-2-Methoxy-*cis*-decalin

The *cis*- form (III and IV) is found both experimentally and by our calculations to have

the larger energy difference between the two conformations (axial and equatorial). This is due to the destabilisation of the axial conformer IV by steric clashes between the methoxy group and the *cis*-fused ring. No such steric clashes occur in the *trans*-form (I and II) and so the energy difference is smaller.

6.6 References to Chapter 6

1. S. Melberg and K. Rasmussen, *J. Mol. Struc.*, **57**, 215 (1979).
2. S. R. Niketic and K. Rasmussen, in *The Consistent Force Field*, Springer, New York (1977).
3. K. Rasmussen, *Potential Energy Functions in Conformational Analysis,* in *Lecture Notes in Chemistry, Vol 37*, Springer-Verlag, Berlin & Heidelberg (1985).
4. M.D. Harmony, V.W. Laurie, R.L. Kuczkowski, R.H. Schwendeman, D.A. Ramsay, F.J. Lovas, W.J. Lafferty, and A.G. Maki, *J. Phys. Chem. Ref. Data.*, **8**, 619 (1979).
5. K. Kimura and K. Kubo, *J. Chem. Phys.*, **30**, 151 (1959).
6. O. Bastiansen, L. Fernholt, H.M. Seip, H. Kambara, and K. Kuchitsu, *J. Mol. Struc.*, **18**, 163 (1973).
7. W. F. Bradford, S. Fitzwater, and L. S. Bartell, *J. Mol. Struc.*, **38**, 185 (1977).
8. T. Iijima, *Bull. Chem. Soc. Jpn.*, **46**, 2311 (1973).
9. T. Iijima, *Bull. Chem. Soc. Jpn.*, **45**, 1291 (1972).
10. R. L. Hilderbrandt and J. D. Wieser, *J. Mol. Struc.*, **15**, 27 (1973).
11. B. Beagley, D.P. Brown, and J.J. Monaghan, *J. Mol. Struc.*, **4**, 233 (1969).
12. H.B. Burgi and L.S. Bartell, *J. Am. Chem. Soc.*, **94**, 5236 (1972).
13. A. Almenningen, O. Bastiansen, and P.N. Skancke, *Acta Chem. Scand.*, **15**, 711 (1961).

14. K. Raghavachari, *J. Chem. Phys.*, **81**, 1383 (1984).

15. A. T. Hagler, P. S. Stern, S. Lifson, and S. Ariel, *J. Am. Chem. Soc.*, **101**, 813 (1979).

16. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case, *J. Comput. Chem.*, **7**, 230 (1986).

17. K. Oyanagi and K. Kuchitsu, *Bull. Chem. Soc. Jpn.*, **51**, 2237 (1978).

18. M. Davis and O. Hassel, *Acta Chem. Scand.*, **17**, 1181 (1963).

19. H.J. Geise, W.J. Adams, and L.S. Bartell, *Tetrahedron*, **25**, 3045 (1969).

20. H.E. Breed, G. Gundersen, and R. Seip, *Acta Chem. Scand.*, **A33**, 225 (1979).

21. U. Burkert, *Tetrahedron*, **35**, 1945 (1979).

22. P.H. Kasai and R.J. Myers, *J. Chem. Phys.*, **30**, 1096 (1959).

23. N.L. Allinger, L. Schafer, K. Siam, and C. Van Alsenoy, *J. Comput. Chem.*, **6**, 331 (1985).

24. R.M. Lees and J.G. Baker, *J. Chem. Phys.*, **48**, 5299 (1968).

25. Y. Sasada, M. Takano, and T. Satoh, *J. Mol. Spect.*, **38**, 33 (1971).

26. A.A. Abdurakhamanov and L. M. Imanov, *Zh. Strukt. Khim.*, **15**, 42 (1974).

27. D.R. Truax and H. Wieser, *Chem. Soc. Rev.*, **5**, 411 (1976).

28. P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, *Proteins: Structure, Function and Genetics*, **4**, 31-47 (1988).

29. T. Shimanouchi, *Tables of Molecular Vibrational Frequencies,* in *National Standards Reference Data Series*, National Bureau of Standards, Washington DC (1972).

30. T.Shimanouchi, H. Matsuura, Y. Ogawa, and I. Harada, *J. Phys. Chem. Reference Data*, **7(4)**, 1323 (1978).

31. K.W. Logan, H.R. Danner, J.D. Gault, and H. Kim, *J. Chem. Phys.*, **59**, 2305 (1973).

32. J. H. Schachtschneider and R. G. Snyder, *Spectrochim. Acta*, **21**, 169 (1965).

33. J.C. Evans and H.J. Bernstein, *Can. J. Chem.*, **34**, 1037 (1956).

34. S. Lifson and A. Warshel, *J. Chem. Phys.*, **49**, 5116 (1968).

35. R.G. Snyder and G. Zerbi, *Spectrochim. Acta*, **23A**, 391 (1967).

36. C.E. Blom and C.L. Altona, *Mol. Phys.*, **34**, 557 (1977).

37. O. Gebhardt and S.J. Cyvin, *J. Mol. Struc.*, **12**, 205 (1972).

38. O.H. Ellestad and P. Klaboe, *Spectrochim. Acta*, **27A**, 1025 (1971).

39. H.M. Pickett and H.L. Strauss, *J. Chem. Phys.*, **53**, 376 (1970).

40. W.R. Ward, *Spectrochim. Acta*, **21**, 1311 (1965).

41. A. J. Barnes and H. E. Hallam, *Trans. Faraday Soc.*, **66**, 1932 (1970).

42. K. Tasaki and A. Abe, *Polymer J.*, **17**, 641 (1985).

43. O. Ermer, *Structure and Bonding*, **27**, 161, Berlin (1976).

44. S. Weiss and G.E. Leroi, *J. Chem. Phys.*, **48**, 962 (1968).

45. D.A.C. Compton, S. Montero, and W.F. Murphy, *J. Phys. Chem.*, **84** , 3587 (1980).

46. A.L. Verma, W.F. Murphy, and H.J. Bernstein, *J. Chem. Phys.*, **60**, 1540 (1974).

47. M.R. Whalon and C.H. Bushweller, *J. Org. Chem.*, **49**, 1185 (1984).

48. P. S. Stern, M. Chorev, M. Goodman, and A. T. Hagler, *Biopolymers*, **22**, 1885 (1983).

49. J. R. Durig and D. A. C. Compton, *J. Chem. Phys.*, **69**, 4713-4719 (1978).

50. T. Kitagawa and T. Miyazawa, *Bull. Chem. Soc. Jpn.*, **41**, 1976 (1968).

51. J.P. Perchard, J.C. Monier, and P. Dizabo, *Spectrochim. Acta*, **27A**, 447 (1971).

52. K. Tasaki, Y. Sasanuma, I. Ando, and A. Abe, *Bull. Chem. Soc. Jpn.*, **57**, 2391 (1984).

53. M. Nakata, Y. Furukawa, H. Hamaguchi, and M. Tasumi, *45th National Meeting of the Chemical Society of Japan* (Tokyo, April 1982, Abstr. No. 4U15).

54. J.R. Durig, S.M. Craven, J.H. Mulligan, and C.W. Hawley, *J. Chem. Phys.*, **58**, 1281 (1973).

55. A.J. Kirby, in *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer, Berlin (1983).

56. G.F. Smits, M.C. Krol, P.N. van Kampen, and C. Altona, *J. Mol. Struc. (Theochem.)*, **139**, 247 (1986).

57. R.K. Kakar and C.R. Quade, *J. Chem. Phys.*, **72**, 4300 (1980).

58. R.K. Kakar and P.J. Seibt, *J. Chem. Phys.*, **57**, 4060 (1972).

59. C. Tanaka, *Nippon Kagaku Zasshi*, **83**, 661 (1962).

60. E.T. Beynon, Jr. and J. McKetta, *J. Phys. Chem.*, **67**, 2761 (1963).

61. J-P. Perchard and M-L. Josien, *J. Chim. Phys. Physiochim. Biol.*, **65**, 1834 (1968).

62. M. Takano, Y. Sasada, and T. Satoh, *J. Mol. Spectrosc.*, **26**, 157-162 (1968).

63. L. Radom, W.J. Hehre, and J.A. Pople, *J. Am. Chem. Soc.*, **94**, 2371 (1972).

64. J. Murto, M. Rasanen, A. Aspiala, and T. Lotta, *J. Mol. Struc.*, **108**, 99 (1984).

65. S. Lifson, A. T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979).

66. H. Booth and J.R. Everett, *J. Chem. Soc., Perkin Trans. II*, 255 (1980).

67. J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometalic Compounds,* Academic Press, New York, NY (1970).

68. J.B. Pedley and J. Rylance, *Sussex N.P.L. Computer Analysed Thermochemical Data: Organic and Organometallic Compounds,* Academic Press, Univ. of Sussex (1977).

69. M. Squillacote, R.S. Sheridan, O.L. Chapman, and F.A.L. Anet, *J. Am. Chem. Soc.*, **97**, 3244 (1975).

70. H. Booth and K.A. Khedhair, *J. Chem. Soc., Chem. Commun.*, 467 (1985).

71. I.G. Mursakulov, E.A. Ramazanov, M.M. Guseinov, N.S. Zefirov, V.V. Samoshin, and E.L. Eliel, *Tetrahedron*, **36**, 1885 (1980).

72. G.G. Engerholm, A.C. Luntz, and W.D. Gwinn, *J. Chem. Phys.*, **50**, 2446 (1969).

73. M. Anteunis, A. Geens, and R. Van Cauwenberghe, *Bull. Soc. Chim. Belg.*, **82**, 573 (1973).

74. D. Tavernier, F. De Pessemier, and M. Anteunis, *Bull. Soc. Chim. Belg.*, **84**, 333 (1975).

75. P. Deslongchamps, in *Stereoelectronic Effects in Organic Chemistry*, Pergamon, Oxford (1983).

76. K.B. Wiberg and M.A. Murcko, *J. Am. Chem. Soc.*, **111**, 4821 (1989).

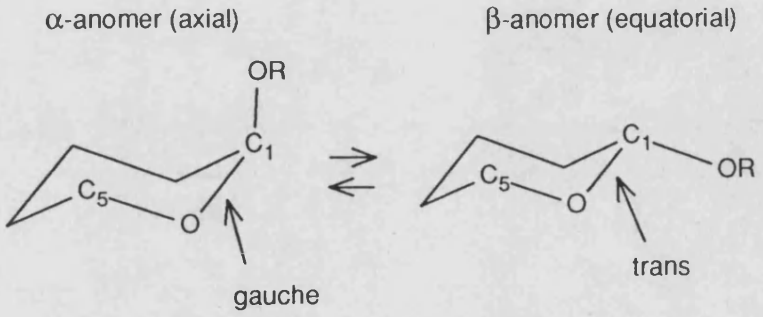
Chapter 7

Modelling the Anomeric Effect

7.1 Introduction

The *anomeric effect* is a conformational effect that was so-named because it was first observed for the anomeric carbon atoms (C1) of pyranose rings.[1] It has since been found to occur for many other types of compounds, but it is important in this context because it has a major influence on the conformation of carbohydrate molecules.

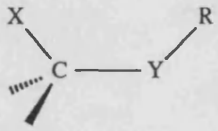
For almost any pyranose derivative, experimental evidence suggests a stabilising of the α anomer (axial C1 substituent) rather than the β anomer (equatorial C1 substituent). This cannot be explained by conventional considerations of conformational analysis.[2]



For substituted cyclohexanes, the preferred conformation is generally the one with the larger number of equatorial substituents, and thus the minimum number of *gauche* interactions and non-bonded repulsions between axial substituents. Similar considerations apply for substituted tetrahydropyrans, which like cyclohexanes, show a preference for equatorial substituents in all cases except one. This is the case where the

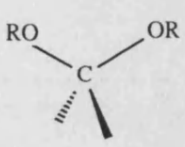
substituent on the tetrahydropyran ring is an electronegative atom (**O,F,Cl**) on one of the carbon atoms adjacent to the ring oxygen. The most abundant molecules showing this pattern of substitution are, of course, the pyranose derivatives discussed above.

The anomeric effect has been interpreted as a preference for a *gauche* conformation about the **C–O–C–O** torsion angle, and occurs also in acyclic **C–O–C–O** systems, which similarly display a preference for *gauche* rotamers. In fact, the anomeric effect has been observed in many different types of compound that contain geminally disubstituted electronegative atoms:[3]

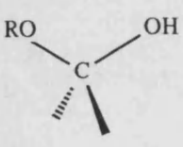


X = O, F, Cl and Y = O or S

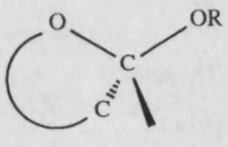
Because our forcefield is at present only seeking to reproduce carbohydrate properties, we shall limit our considerations to those classes of compound for which **X=Y=O**; the acetals and hemiacetals.



Acetal



Hemiacetal

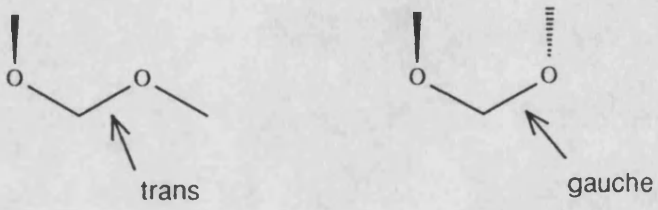


Cyclic Acetal

Pyranose and furanose derivatives, the building blocks of carbohydrate molecules, are examples of cyclic acetals and hemiacetals.

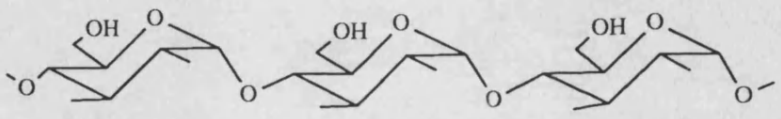
A compound that has commonly been used as a model for the anomeric effect is the simplest acetal, dimethoxymethane.[4-7] Experimental data for this compound can be used to give an idea of the magnitude of the anomeric effect, in terms of conformational energy. Dimethoxymethane has long been known to favour the *gauche,gauche* conformation,[8] with a *trans-gauche* energy difference for each **C–O** bond of about 1.7 kcal/mol.[8]

Dimethoxymethane



Rel. Energy (kcal/mol)	1.7	0

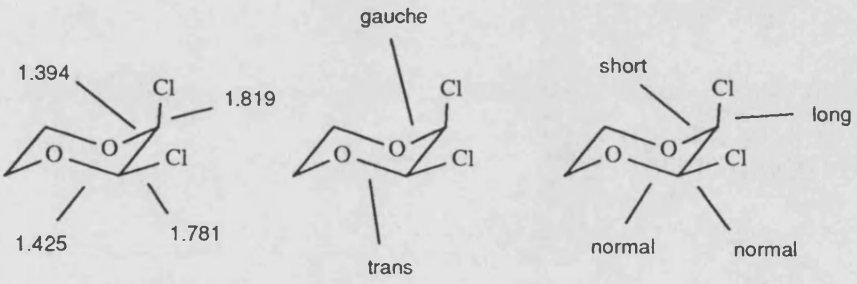
However, an isolated **C–O** bond, such as that in ethylmethylether, prefers to be in the *trans* conformation by about 1.5 kcal/mol[9] (see Chapter 6). The anomeric effect can therefore be seen, from the difference between these values, to account for something in the region of 3 kcal/mol in favour of the *gauche* rotamer. This is a large amount of energy in conformational terms, especially since it is dependent on the rotation about a single bond. The importance of the anomeric effect becomes increasingly clear when we consider that the majority of oligo- and polysaccharides consist of pyranose rings linked through acetal **C–O** bonds.



1,4 - linked saccharide chain

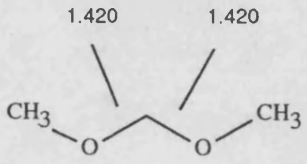
Although the clearest manifestation of the anomeric effect is the preference for *gauche* conformations, detailed examination of the geometries shows that the **C–O** bond lengths are also affected. It has been known for some time that when there are more than one electronegative atoms (**X**) on a given carbon, the **C–X** bonds are shorter than for the monosubstituted case.[10] This has been attributed to an 'electronegativity effect',[4] which is described as follows. Two or more electronegative substituents (**X**) on a carbon atom will lead to a larger partial positive charge on that carbon; the **C–X** bonds will therefore be shortened because of the increased electrostatic attraction between the **C** and **X** atoms. However, there is experimental evidence to suggest that an additional factor is at work in compounds exhibiting an anomeric effect, as bond lengths are found to be *dependent* on the values of their torsion angles; an observation that could not be explained by the electronegativity effect.

The torsion angle dependence of bond length has been observed for carbohydrates,[11,12] but perhaps the clearest demonstration of it is seen in the crystal structure of a non-carbohydrate compound, *cis*-2,3-dichloro-1,4-dioxane.[13]

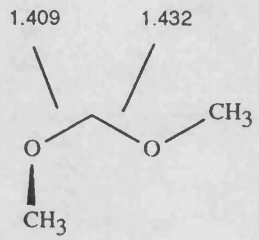


This molecule contains both a *gauche* **C–O–C–Cl** fragment (with the **Cl** axial) and a *trans* fragment (with the **Cl** equatorial). The crystal structure shows the **C–O** and **C–Cl** bonds of the *trans* fragment to be of normal length for alkyl ethers and chlorides, whereas those of the *gauche* fragment demonstrate bond shortening for the **C–O** bond, and bond lengthening for the **C–Cl** bond. The value of the **C–O–C–Cl**

torsion is therefore shown to have a critical effect on the values of the bond lengths. A similar pattern of behaviour is reproduced in *ab initio* calculations of acetals and hemiacetals.[14,15] The bond lengths shown below were obtained by Van Alsenoy *et al.* for dimethoxymethane (using the 4-21G basis set) and show a similar bond shortening/lengthening effect when one of the **C–O–C–O** torsions is *gauche*.



trans,trans



gauche,trans

There are therefore two observations concerning the anomeric effect that must be taken into account when devising a valence forcefield model to account for it. Firstly, the forcefield should be able to reproduce the conformational energy differences of the anomeric fragment, and secondly, it should be able to reproduce the dependency of the bond lengths on torsion angles. In order to construct a forcefield model, however, we must first understand the electronic basis of the anomeric effect.

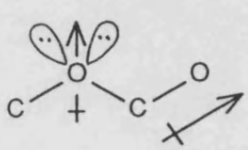
7.2 The Mechanism of the Anomeric Effect

A number of rationalisations of the anomeric effect have been advanced in the literature.[16-18] These have all sought to account, at least qualitatively, for the preference for *gauche* (and axial) conformations: however, most do not predict the bond length changes which are characteristic of systems demonstrating the anomeric effect. The theoretical work by Pople, Jeffrey and co-workers on anomeric model compounds[19,20,11] has shown that *two* of the previously suggested explanations were substantially correct, and that *both* contributed to the complete description of the

anomeric effect. We shall refer to these contributing effects as *(i)* dipole-dipole repulsion, and *(ii)* n-σ^* conjugation.

7.2.1 Dipole-Dipole Repulsion

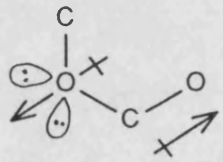
This was one of the first explanations proposed for the anomeric effect,[16] and seeks to account only for the conformational preferences. The theory suggests that the *trans* rotamer about a **C–O** bond in an acetal is destabilised by an electrostatic repulsion between dipoles. These dipoles are the bond dipole of one **C–O** bond, and a local dipole on the adjacent oxygen, caused by the presence of the lone-pair electrons.



trans

dipoles aligned

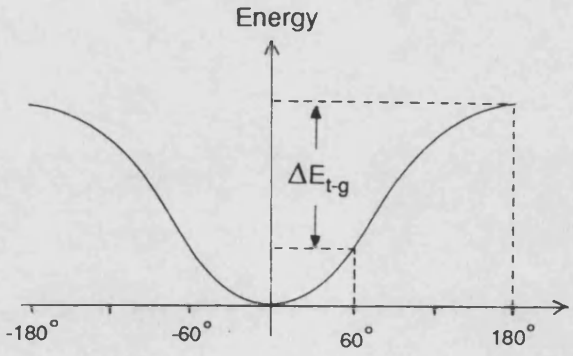
max. repulsion



cis

dipoles opposed

min. repulsion

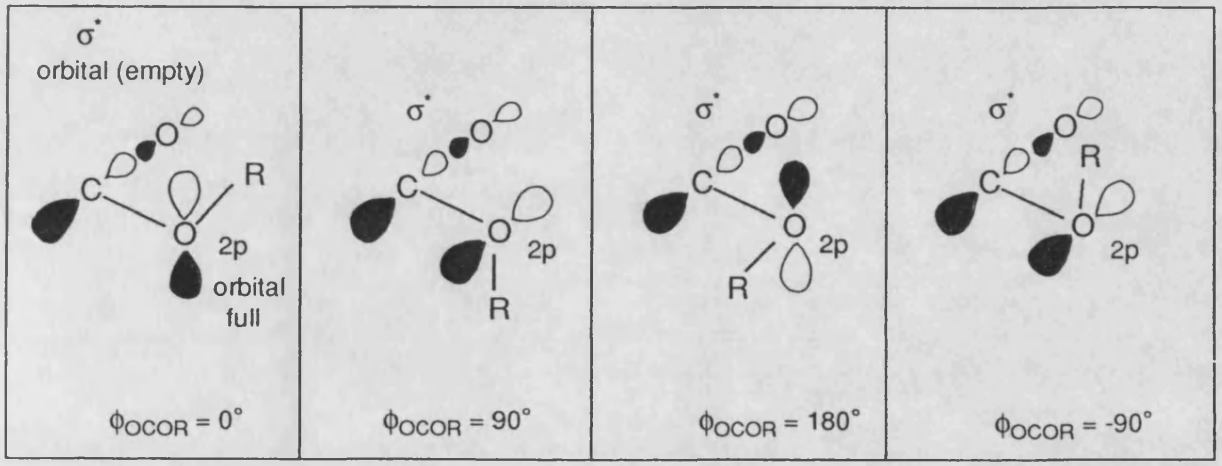


The maximum repulsion will be when the dipoles are aligned (at $\phi_{OCOC} = 180^\circ$, *trans*) and will be at a minimum when the dipoles are opposed ($\phi_{OCOC} = 0^\circ$, *cis*). The *trans*

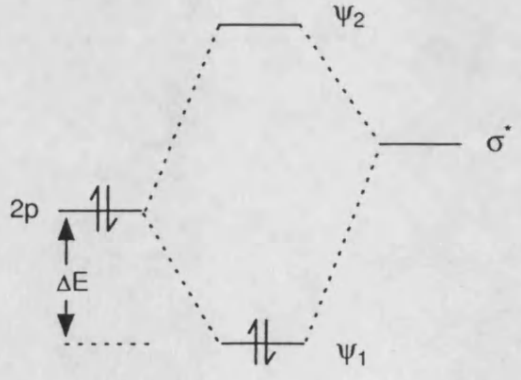
minimum will therefore be destabilised with respect to the *gauche*. This mechanism, although regarded as being a contributor to the anomeric effect,[2] does not account for the variation of bond length with torsion angle.

7.2.2 n-σ Conjugation*

n-σ* Conjugation was first suggested by Lucken[18] and is a localised molecular orbital description of the anomeric effect. At certain values of the anomeric torsion angle, orbital overlap occurs between a lone-pair of electrons in an orbital on one of the oxygen atoms, and a σ* antibonding orbital of the adjacent **C–O** bond.

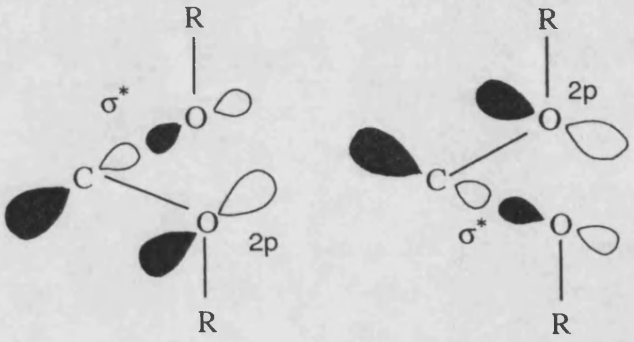


As can be seen from the diagram above, the oxygen lone-pair (in a 2p orbital) has the correct orientation for overlap with the σ* orbital at ϕ values of near +90° and -90°. This overlap lowers the energy of the lone-pair electrons (as shown in the energy diagram below) and thus stabilises conformations for which it can occur.



The observed bond length changes are also explained by this mechanism: the n-σ^* overlap gives a π-type interaction between the 'donor' oxygen and the carbon, resulting in a stronger, *shorter* bond. Conversely, n-σ^* overlap *lengthens* the other **C–O** bond since electron density in its antibonding orbital will be increased. When no n-σ^* overlap occurs, however, the bond lengths will not deviate from their natural values. *Gauche* conformations have ϕ values of about 60°, sufficiently close to 90° for n-σ^* overlap to occur, and so they will be stabilised with respect to the *trans* conformation (for which ϕ = 180°).

In acetals and hemiacetals, two such orbital overlaps can occur simultaneously, if both **C–O** torsion angles are at (or near) 90°:

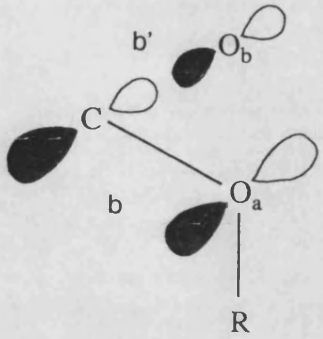


This explains why the *gauche,gauche* conformer of dimethoxymethane is the most stable.[8]

This mechanism is consistent with both the experimental and *ab initio* results regarding bond lengths (see Section 7.1). A summary of the effects resulting from n-σ^* conjugation is given in Table 7.1.

Table 7.1 Effects of n-σ^* Conjugation

ϕ (**R–O_a–C–O_b**)	**C–O_a** (b)	**C–O_b** (b')	Relative Energy
0°	normal	normal	high
90°	short	long	low
180°	normal	normal	high
270° (=-90°)	short	long	low
360° (=0°)	normal	normal	high



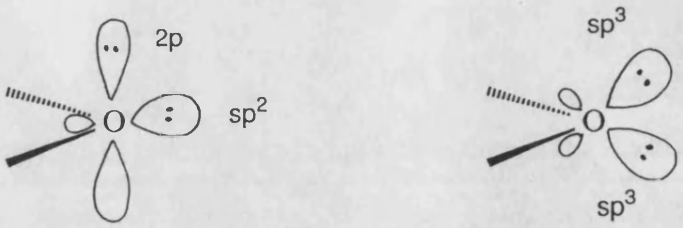
n-σ^* Conjugation is sometimes described as 'double-bond no-bond resonance',[3] as the following resonance diagram can be drawn:

However, from diagrams like these, the dependence of this 'resonance' on the rotation about the **R–O–C–O** torsion is not obvious.

Hybridisation of the Lone-Pair Electrons

The description given above for n-σ^* conjugation depicted the lone-pair electrons on the 'donor' oxygen (**O_a**) as occupying a 2p orbital. This requires the oxygen atom **O_a** to be sp^2 hybridised. The second lone-pair on **O_a**, not involved in n-σ^* conjugation, is therefore in an sp^2 orbital. An alternative way of describing the oxygen lone-pairs is

as two energetically equivalent, sp^3 hybridised orbitals.



n-σ^* Conjugation has been rationalised using both these depictions of hybridisation,[21] and Kirby has argued that the two representations are mathematically almost equivalent.[2] However, recent evidence from a statistical analysis of crystallographic data supports the sp^2 hybridised oxygen, with non-equivalent lone-pairs.[22] Photo-electron spectroscopy also suggests non-equivalence of lone-pairs, as two different ionisation energies are found for the lone-pair electrons.[23,24]

7.3 Previous Empirical Approaches to the Anomeric Effect

Before considering how to adapt our forcefield to reproduce the anomeric effect, it is relevant to consider the methods that have been used to account for it in previous computational studies.

Rigid Geometry Calculations

The earliest computational modelling of carbohydrate molecules was performed using a rigid geometry method which considered only non-bond interactions (the so-called 'Hard-Sphere' method).[25] However, because the anomeric effect is so important in determining the conformations about glycosidic linkages, a further 'exo-anomeric' energy term was added that gives an additional energy contribution dependent solely on the glycosidic torsion angle.

$$E_{Total} = E_{non-bond} + E$$

$$E_{Total} = E_{non-bond} + E_{exo-anomeric}$$

The function used to describe this exo-anomeric energy is a three component Fourier series.

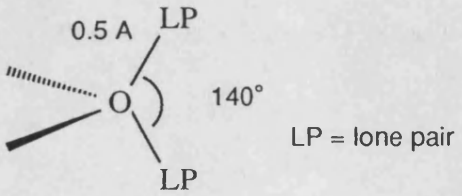
$$E_{exo-anomeric} = V_1(1-\cos\phi) + V_2(1-\cos 2\phi) + V_3(1-\cos 3\phi) + k$$

The parameters of this function were determined by fitting to the torsional rotation potential of a model compound, dimethoxymethane, found by *ab initio* calculations.[26] This has become known as the HSEA (Hard-Sphere Exo-Anomeric) method, and has been used in conjunction with NMR in the study of oligosaccharide conformation.[26,27]

Flexible Geometry Calculations

To our knowledge, the only flexible geometry forcefields that have specifically attempted to reproduce the anomeric effect are those of Allinger, MM1 and MM2.[28,4]

These forcefields include lone-pairs explicitly (as pseudo-atoms, **LP**) situated 0.5 Å from the oxygen atom bearing them, and with a **LP–O–LP** 'valence angle' of 140°. The **LP–O** bonds are also considered to have an associated dipole moment (with the lone-pair negative). An advantage of this depiction of lone-pairs is that the dipole-dipole repulsion (discussed in Section 7.2.1) will be accounted for without further modification of the forcefield.



The MM1 and MM2 forcefields further seek to reproduce the energetics of n-σ^* conjugation by using a twofold torsional term for the **LP–O–C–O** torsion angle.[28,4]

$$E_{OCOLP} = \frac{V_2}{2}(1 - \cos 2\phi_{OCOLP})$$

This gives a minimum value of E_{OCOLP} at ϕ values of 0° and 180°, and a maximum at 90° and -90°. The justification made for this functional form is that the n-σ^* overlap will be greatest when the **O–C** and **O–LP** bonds of the **O–C–O–LP** fragment are aligned.[28]

So far, the calculational methods described have attempted to account only for the energetics of the anomeric effect. In 1984, however, Allinger and Norskov-Lauritsen[4] described an amendment to MM2 aimed at reproducing the accompanying bond length changes. Although they acknowledged that a bond-torsion cross term would be the most straightforward way to account for the dependence of a bond length on a torsion angle, they did not use this approach because of the difficulties involved in implementing such a term in the MM2 program. Instead, they used an approximation that amounts to a torsionally dependent b_0 parameter. (Although Allinger uses the symbol l_0 for his parameter, it is essentially the same quantity as our b_0 parameter, i.e. the 'strain-free' bond length.)

The functional form that relates l_0 to the torsion angles within the acetal fragment (**C_1–O_2–C_3–O_4–C_5**) was selected to reproduce both bond shortening and lengthening at the appropriate torsion angle values. For example, the normal value of l_0 for the **O_2–C_3** bond is altered by an amount δl, to give a new strain-free bond length l'_0, in the following manner:

$$l'_0 = l_0 + \delta l$$

where

$$\delta l = \frac{k}{2}(1 + \cos 2\phi_{2,3}) - \frac{ck}{2}(1 + \cos 2\phi_{3,4}) + d$$

The results obtained by this method for dimethoxymethane are compared for those for our forcefield in Section 7.6.1.

7.4 A Valence Forcefield Model of the Anomeric Effect

One of the basic philosophies behind the valence forcefield method is that the conformational energy of a molecule can be partitioned into energy terms that have some physical significance. In theory, conformational energy could be represented by any arbitrary function of the atomic positions, so long as it contained a sufficient number of parameters. However, because chemists generally try to understand conformational energy in terms of bond stretches, steric clashes and so forth, functional forms are invariably selected to reflect these perceptions.

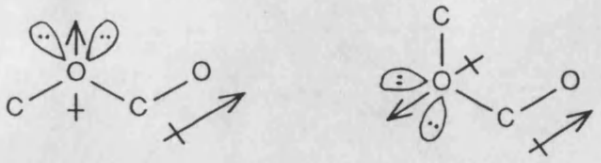
In choosing a forcefield description of the anomeric effect, we have sought to apply this same philosophy, firstly by understanding the underlying electronic mechanisms, and secondly devising simple functional forms that bear an obvious relationship to these mechanisms.

In Section 7.2, two mechanisms that contribute to the anomeric effect were described. These were dipole-dipole repulsion and n-σ^* conjugation. We shall now show how each of these may be accounted for in a valence forcefield formalism.

7.4.1 Dipole-Dipole Repulsion

As described in Section 7.2.1, this is the repulsion that occurs between a local dipole

on an oxygen atom and that of the adjacent **C–O** bond.



trans

dipoles aligned

max. repulsion

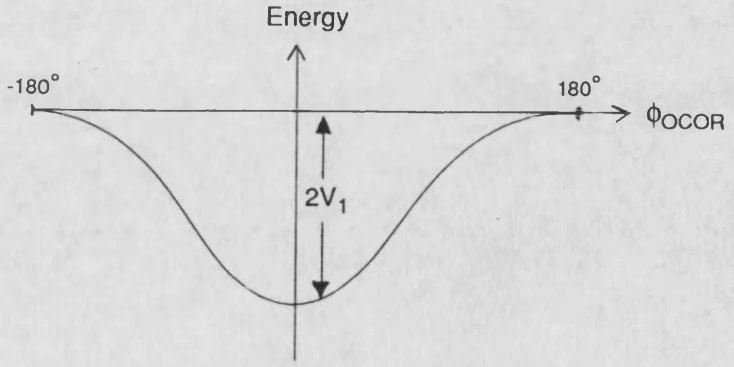
cis

dipoles opposed

min. repulsion

This type of electrostatic repulsion was not accounted for in our original forcefield, as oxygen lone-pairs are not explicitly included. It was a relatively easy matter, however, to incorporate this effect into the forcefield by the use of the onefold (V_1) torsional term. From the diagram above, the dipole-dipole repulsion will be at a maximum when the **O–C–O–R** fragment is *trans* ($\phi = 180^\circ$) and will fall to a minimum in the *cis* position ($\phi = 0^\circ$). By using a negative V_1 parameter (cf. dimethoxyethane, Section 6.4.2) this energy difference can be reproduced.

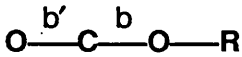
$$E = V_1(1 - \cos\phi_{OCOR})$$



7.4.2 n-σ^* Conjugation

This localised molecular orbital description of the anomeric effect was described in Section 7.2.2. Conformations of the **O–C–O–R** fragment were considered which indicated energy stabilisation and bond length changes to occur at ϕ_{OCOR} values of +90° and -90°. Table 7.1 (see Section 7.2.2) gave a summary of the effects of n-σ^* conjugation. This table can also be considered to summarise the *requirements* of a force-field description of n-σ^* conjugation: a continuous function is needed that reproduces the effects shown in Table 7.1.

The quantities b and b' represent the central and terminal **C–O** bond lengths in the **O–C–O–R** fragment (see Table 7.1). As both b and b' are dependent on ϕ_{OCOR}, the most appropriate way in which to account for this was considered to be with two bond-torsion cross terms. These cross terms firstly relate ϕ_{OCOR} and b, and secondly ϕ_{OCOR} and b'.



$$E_{b\phi} = K_{b\phi}\,(b - b_0)\,(1 - \cos 2\phi_{OCOR}) \quad \text{bond shortening term}$$

$$E_{b'\phi} = -K_{b'\phi}\,(b' - b'_0)\,(1 - \cos 2\phi_{OCOR}) \quad \text{bond lengthening term}$$

The functional forms of these two terms are almost identical except for the presence of a minus sign in the second term. This is what leads to the bond *lengthening* of b', rather than the bond *shortening* of b, and the reason for this will become clearer below.

The way in which these functional forms serve to reproduce the effects shown in Table 7.1 can be seen by plotting the function as a three-dimensional energy surface. Figure 7.1 shows how the energy $E_{b\phi}$ varies for the b/ϕ cross term. The low energy areas of the surface can be seen to occur at values of ϕ of +90° and -90°, and at negative values of Δb ($=b - b_0$). This should therefore reproduce *both* the stabilisation

Figure 7.1 Potential Surface for the b/ϕ Cross Term

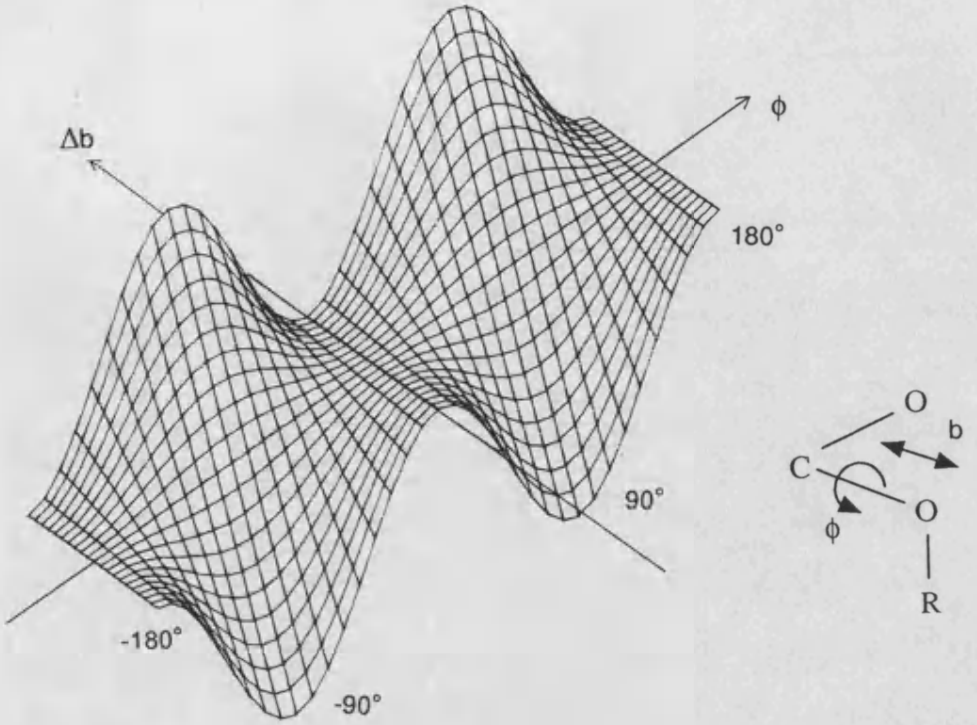
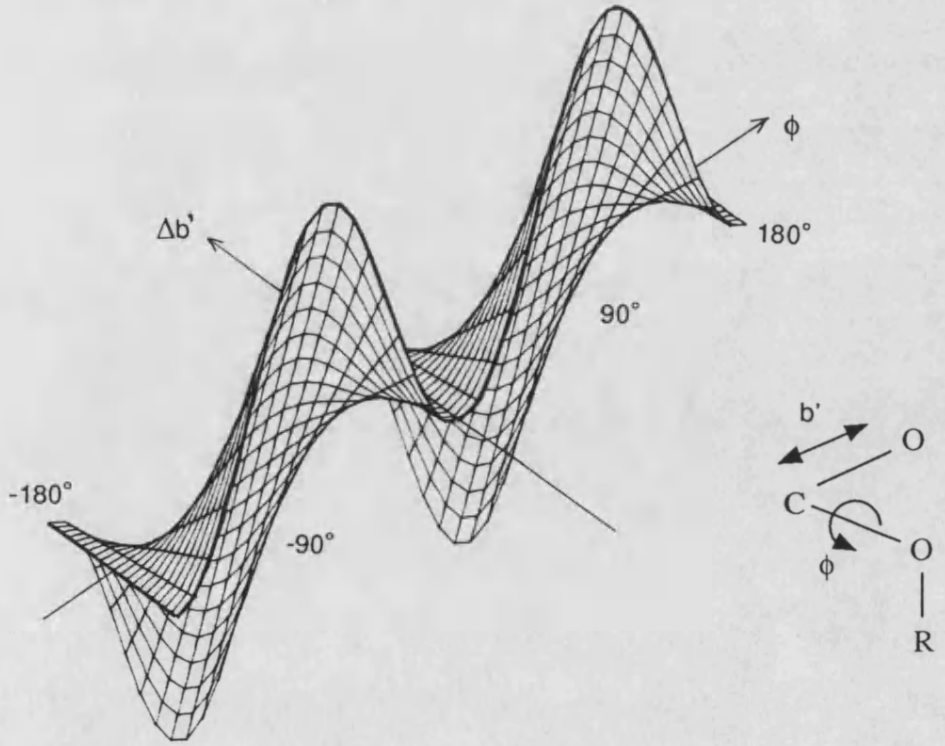


Figure 7.2 Potential Surface for the b'/ϕ Cross Term



energy *and* the short bond length required at these torsion angles. At ϕ values of 0° and 180°, where no orbital overlap occurs, the energy is zero (as $1-\cos\phi = 0$) and there is no effect on bond length.

A similar plot of the b'/ϕ cross term is shown in Figure 7.2. As expected from their functional forms, the sign of the energy is now *reversed* and the low energy areas of the surface occur this time at positive $\Delta b'$ values. This therefore results in a bond lengthening effect.

7.5 The Bond-Torsion Potential Surface

In order to further clarify our choice for the functional forms discussed in Section 7.4.2, it is worthwhile illustrating how these energy terms combine with the other terms in the forcefield to give the total potential surface.

We shall consider only the bond-shortening term here, but because of the similarity of the two terms, the application of these arguments to the bond-lengthening term is directly analogous.

The energy terms present in the forcefield that relate directly to the values of b and ϕ are the bond stretch term, the torsion term and the new bond-torsion cross term.

$$E_{Total} = E_{bond} + E_{b\phi} + E_{torsion}$$

The potential energy surface for $E_{b\phi}$ was presented in the previous section (Figure 7.1). We will now look at the potential surfaces that are given by the other energy terms in the above expression, E_{bond} and $E_{torsion}$.

The Bond Stretch Energy (E_{bond})

From the forcefield equation (3.1) the bond stretch energy is given by a Morse potential:

$$E_{bond} = K_b\left[1 - \exp(-\alpha\{b - b_0\})\right]^2 - K_b$$

A plot of the energy surface resulting from this function is shown in Figure 7.3, showing the surface to be 'valley' shaped (this holds for b values close to b_0).[29] The energy is independent of ϕ (as expected) but rises steeply when the bond is stretched (Δb positive) or compressed (Δb negative).

The Torsion Energy ($E_{torsion}$)

Again, from the forcefield equation (3.1), the torsional energy was given as the sum of three cosine terms.

$$E_{torsion} = V_1(1 + \cos\phi) + V_2(1 - \cos 2\phi) + V_3(1 + \cos 3\phi)$$

The V_2 parameter for the **O–C–O–R** torsion is zero, and so this leaves only the onefold and threefold term to describe the torsional energy. The onefold term was shown in Section 7.4.1 to mimic the effect of dipole-dipole repulsion, as long as V_1 has a negative value. The threefold term derives from the simple preference for staggered conformations (ϕ = 60°, 180°, -60°) over eclipsed (ϕ = 0°, 120°, -120°). The energy surfaces for these two torsional terms are shown in Figure 7.4(a) & (b), and the total torsional energy, $E_{torsion}$, given by the sum of these two terms, is shown in Figure 7.4(c).

Figure 7.3 Potential Surface for the Bond-Stretch Term

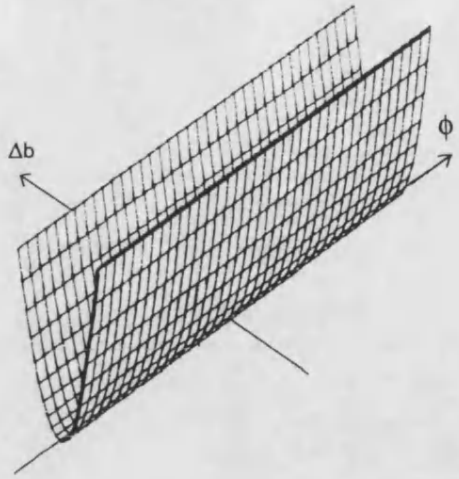


Figure 7.4 Potential Surface for the Onefold and Threefold Torsion Terms

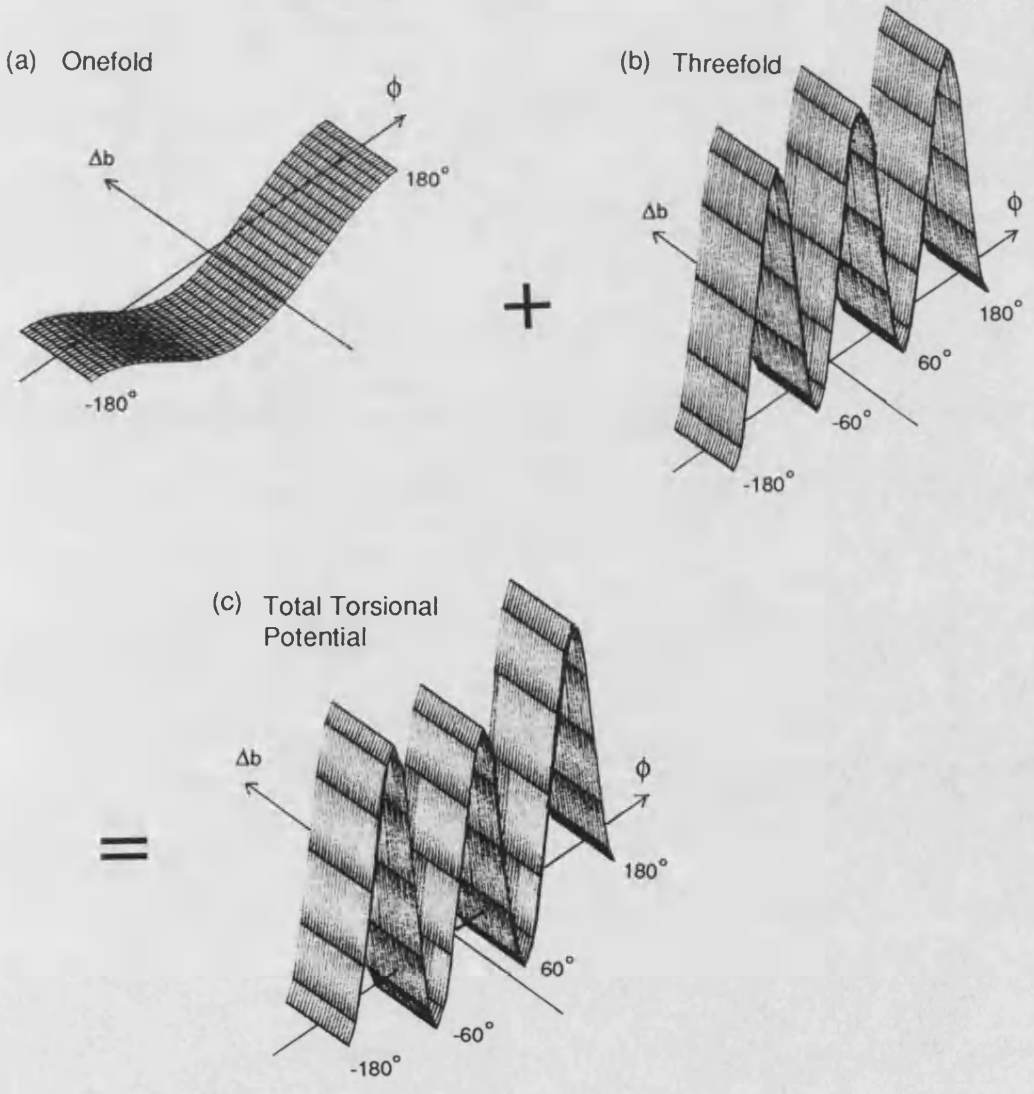


Figure 7.5 Summation of the b/ϕ Cross Term and the Bond-Stretch Terms

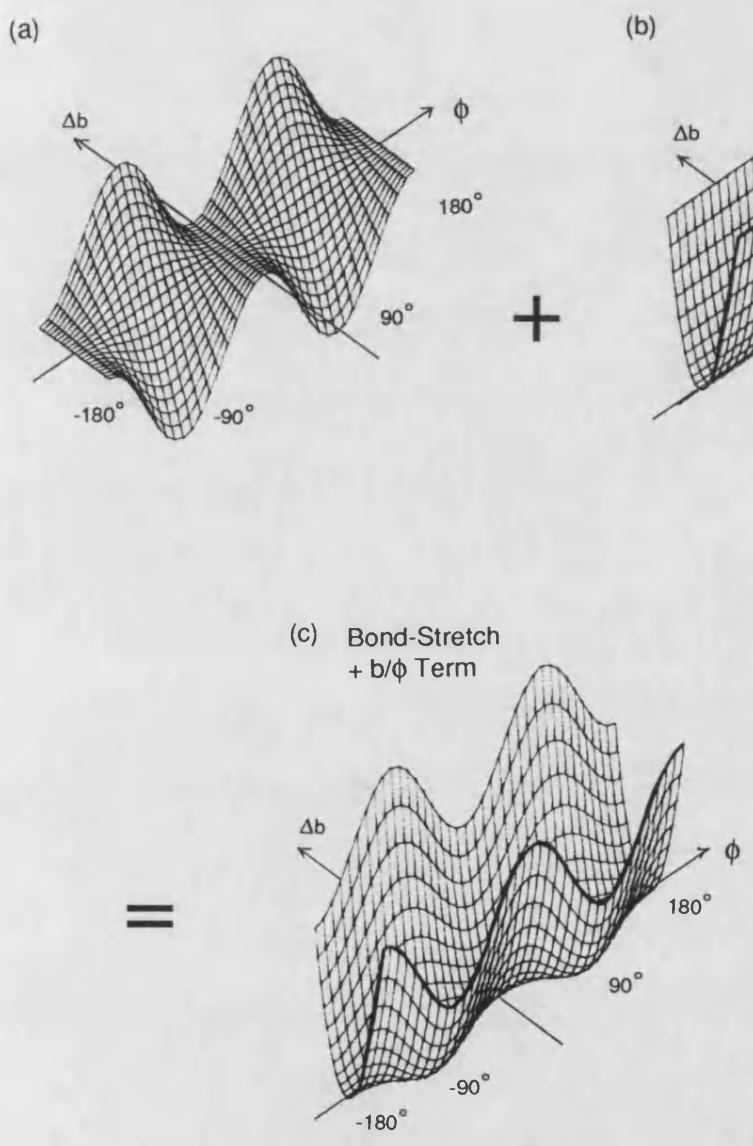
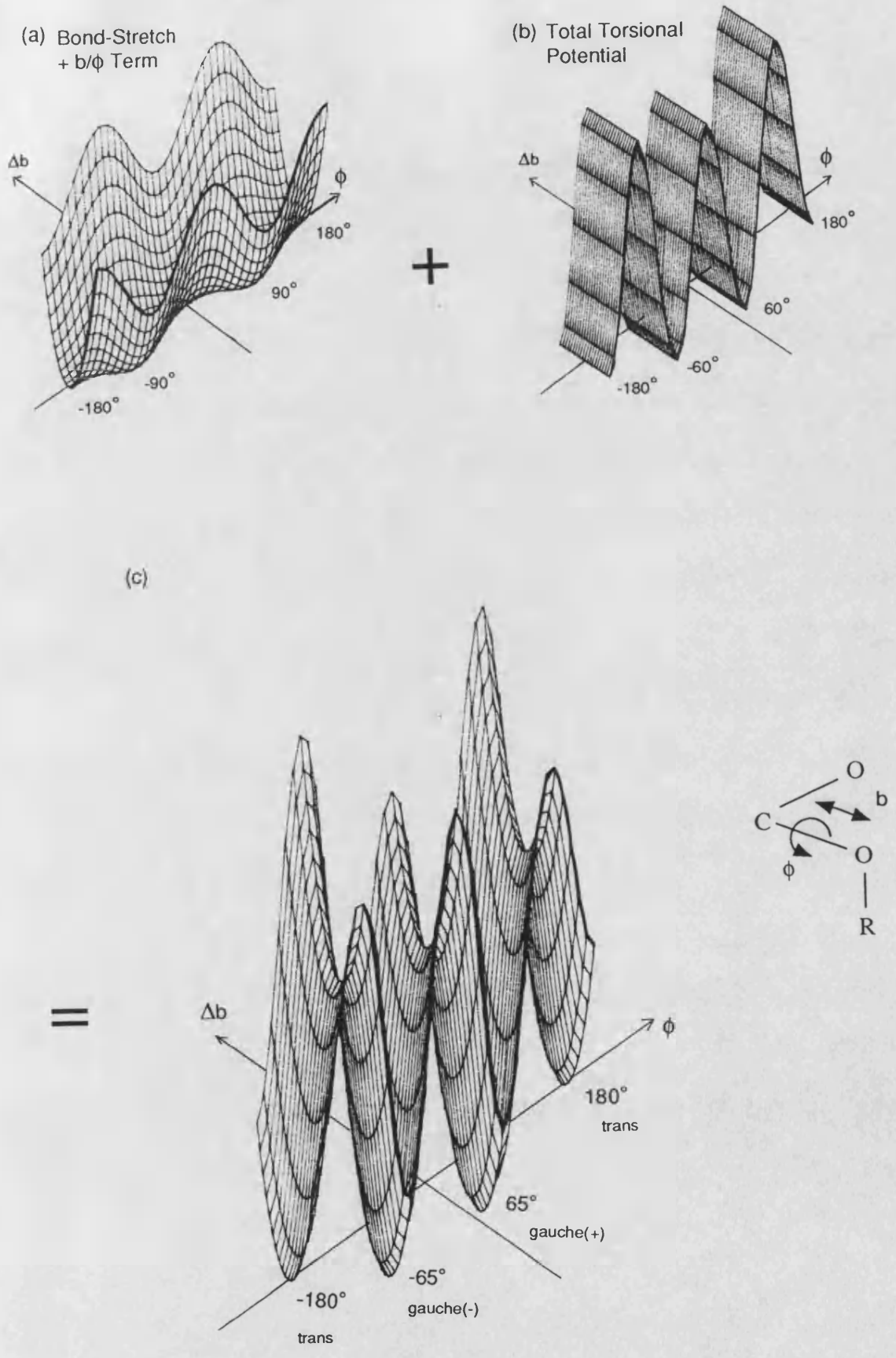


Figure 7.6 Summation to give the Total Bond-Torsion Potential Surface



The Total Bond-Torsion Energy (E_{Total})

We are now in a position to construct the *total* bond-torsion potential surface, by summing the component surfaces for E_{bond}, $E_{b\phi}$ and $E_{torsion}$. This summation is illustrated in Figures 7.5 and 7.6.

The total bond-torsion potential energy surface shown in Figure 7.6(c) should now account for the behaviour of the anomeric bonds. In Figure 7.6(c), three minima are apparent, corresponding to the *trans*, *gauche(+)* and *gauche(-)* conformations. It should be noted that the two *gauche* minima are at lower energy than the *trans*, thus reflecting the experimentally observed preference for the *gauche* conformations. Additionally, the *gauche* minima are shown to lie *in front of the* ϕ *axis* in Figure 7.6(c): this results from the bond-shortening effect of the bond-torsion cross term.

It seems, therefore, that the functional forms used here (i.e. the onefold term to account for dipole-dipole repulsion, and the bond-torsion term to account for n-σ^* conjugation) should, at least from these qualitative considerations, be able to reproduce the observations caused by the anomeric effect.

Note that although we have only considered the bond-shortening term here (b/ϕ); the bond-lengthening term (b'/ϕ) will give a similar energy surface to Figure 7.6(c), except that the *gauche* minima will now lie *behind* the ϕ axis, indicating bond-lengthening to be favoured.

The remainder of this chapter deals with the parameterisation of these functions, and the application of the resulting forcefield to the study of acetal model compounds.

7.6 Determination of the Anomeric Parameters

Having decided on the functional form to be used to represent the anomeric effect, it remains to determine the values of the parameters for these new functions, as well as the more conventional parameters for the **O–C–O** unit.

Table 7.2 Anomeric Parameters

Unit[a]	Parameter	Observables
O–C_a O–C–O C–O–C–O	b_0 θ_0 $K_{b\phi}$,$K_{b'\phi}$	Geometry of Dimethoxymethane
O–C_a O–C–O H–(C–O)–O O–(C–H)–O C–O–C–O	K_b K_θ, $K_{bb'}$, $K_{b\theta}$ $K_{\theta\theta'}$ $K_{\theta\theta'}$ V_3, $K_{\theta\theta'\phi}$	Vibrational Frequencies of 1,3,5-Trioxane
C–O–C–O	V_1	Conformational Energies of Dimethoxymethane

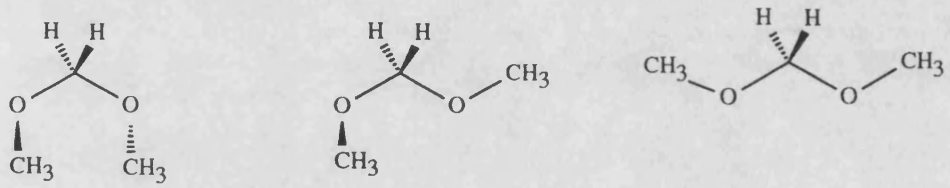
a C_a = anomeric carbon atom (see Appendix I for a description of atom types)

We decided to use acetals (rather than hemiacetals) as model compounds for the anomeric effect for two reasons. Firstly, very few gas phase studies have been made for hemiacetals (perhaps because they are less volatile due to hydrogen bonding); and secondly, the very large 1,4 electrostatic attraction occurring in the $^{\delta-}$O–C–O–H$^{\delta+}$ fragment could obscure the more subtle nature of the anomeric effect itself.

Table 7.2 shows the parameters relating to the anomeric fragment, and the data that they were fitted to. (The final parameter values are included in the tables of forcefield parameters in Appendix I.)

7.6.1 Results for Dimethoxymethane

In so far as is possible, we have attempted to fit forcefield parameters to gas phase experimental data. Unfortunately, good quality gas phase structural data for acetals (electron diffraction or microwave) has proved to be very scarce. The r_a geometry for dimethoxymethane has been determined by Astrup[30] and showed it to be predominantly in the *gauche,gauche* conformation.



gauche,gauche gauche,trans trans,trans

Conformations of Dimethoxymethane

However, some of the structural features found in this determination - particularly the **C–O** bond lengths - are the subject of some controversy.[15,4] From her analysis of the radial distribution function given by electron diffraction, Astrup obtained very different values for the **CH_3–O** and **CH_2–O** bond lengths (1.432 Å and 1.382 Å respectively) which are not supported by *ab initio* calculations.[15,7] A single peak in the radial distribution function was found for the **C–O** bonds of dimethoxymethane, and this may indeed correspond to two different bond lengths: however, Van Alsenoy *et al.* argue that distances within 0.05 Å cannot be resolved with confidence from the radial distribution function, and that therefore only the *average* **C–O** bond length may be taken as an observable.

In addition to Astrup's data therefore, we have decided to consider the fully optimised *ab initio* geometries of Van Alsenoy and co-workers.[15] Although using theoretical data rather than experimental data runs contrary to the original philosophy of the Consistent Forcefield concept,[31] the use of these particular geometries does have

some justification. Van Alsenoy and his co-workers have carried out extensive comparisons of their *ab initio* calculations with those of electron diffraction measurements; [32] and this enables the estimation of 'r_g' values from *ab initio* results. The comparison showed **C–O** bond lengths determined by *ab initio* calculations (4-21G basis set) to be between 0.019 Å and 0.026 Å longer than corresponding r_g values. We have therefore followed the example of Allinger (who also used Van Alsenoy's geometries for dimethoxymethane to parameterise the forcefield MM2(82)[4]) and corrected the *ab initio* bond lengths by -0.023 Å (an average of -0.019 Å and -0.026 Å).

Four geometries for Dimethoxymethane are shown in Table 7.3, and these relate to: (i) the experimental r_a values, (ii) the 'r_g-corrected' *ab initio* values, (iii) the results obtained with our forcefield, and (iv) the geometry obtained from MM2(82). We fitted three of our forcefield parameters to the *ab initio* **C**H_2**–O** bond length values (emboldened) in Table 7.3: these parameters were b_0 for the **O–C**$_a$ bond, and the $K_{b\phi}$ and $K_{b'\phi}$ parameters for the **C–O–C–O** torsion. As can be seen from the table, the bond lengths from our forcefield reproduce those of the *ab initio* geometry almost exactly. The MM2(82) geometry is taken from reference 4, and also seems to have been fitted to the same corrected *ab initio* data; again using three similar parameters (k,c and d - see Section 7.3). The results of the two forcefields are thus directly comparable, and it can be seen that the MM2(82) values for the **C**H_2**–O** bond lengths deviate by between 0.012 and 0.016 Å from the corrected *ab initio* bond lengths, compared with a maximum deviation of 0.002 Å from our forcefield. It therefore seems that our bond-torsion cross terms give a better representation of the anomeric effect than the MM2(82) approximation of a torsionally dependent l_0 parameter (see Section 7.3).

Other aspects of the geometry are also worthy of note. In particular, the **C–O–C–O** torsion angle calculated by our forcefield is in excellent agreement with those found both by experiment and *ab initio*. The **O–C–O** angle is indicated by the *ab initio* calculations to be heavily dependent on the conformation, and the two

Table 7.3 Geometries of Dimethoxymethane[a]

Internal	Source	*gauche,gauche*	*gauche,trans*	*trans,trans*
CH_2–O	E. Diff.[b] (r_a)	1.382	-	-
	Ab Initio[c] (4-21G)	**1.399**	**1.386/1.409**	**1.397**
	This work	1.399	1.388/1.411	1.396
	MM2(82)[d]	1.413	1.398/1.425	1.410
CH_3–O	E. Diff.[b] (r_a)	1.432	-	-
	Ab Initio[c] (4-21G)	1.426	1.425/1.419	1.421
	This work	1.425	1.426/1.422	1.423
	MM2(82)[d]	1.422	1.422/1.421	1.422
C–O–C	E. Diff.[b] (r_a)	114.6	-	-
	Ab Initio[c] (4-21G)	114.5	114.9/114.3	114.0
	This work	116.1	116.8/114.8	115.7
	MM2(82)[d]	112.8	112.8/111.9	119.0
O–C–O	E. Diff.[b] (r_a)	114.3	-	-
	Ab Initio[c] (4-21G)	112.4	109.5	105.9
	This work	113.9	109.8	104.2
	MM2(82)[d]	111.7	109.3	106.4
C–O–C–O	E. Diff.[b] (r_a)	63.3	-	-
	Ab Initio[c] (4-21G)	62.4	57.4/179.4	180
	This work	64.0	64.9/178.0	180
	MM2(82)[d]	72.9	73.0/175.0	180

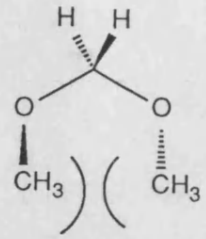
a Bond lengths are in Å, Bond angles in degrees.

b From Reference 30

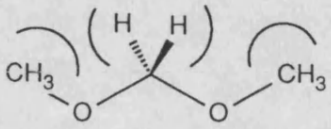
c From Reference 15 (bond lengths corrected by -0.023 Å - see text).

d From Reference 4

forcefield methods also reflect this. From our calculations, this is found to be mainly the result of non-bond repulsions: in the *gauche,gauche* conformation, repulsion between the terminal methyl groups serves to open the **O–C–O** angle, while in the *trans,trans* form, methyl-methylene repulsions cause it to close.



gauche,gauche



trans,trans

Apart from the geometry, the relative energies of the conformations of dimethoxymethane are also important. The only experimental study of the energetics of this molecule was made by Uchida and Kubo in 1956.[8] They measured the temperature dependence of the dipole moment of dimethoxymethane in the gaseous state. The dipole moment of the dimethoxymethane molecule changes with conformation, and increases in the order *gauche,gauche*, *gauche,trans* and *trans,trans*. By making the assumption that the energy difference between the *gauche,gauche* and *gauche,trans* forms was the same as that between the *gauche,trans* and *trans,trans* forms; Uchida and Kubo obtained relative energies of 0, 1.71 and 3.42 kcal/mol respectively for the *gauche,gauche*, *gauche,trans* and *trans,trans* conformers. This assumption was made on the supposition that the *gauche-trans* energy difference for each **C**H_2**–O** bond would be independent of the rotation about the other **C**H_2**–O** bond.

We used these relative energies to fit the V_1 parameter for the **C–O–C–O** torsion angle, which is used to reproduce the effects of dipole-dipole repulsion (as discussed in Section 7.4.1). The value of V_1 that gave the closest fit to these energies was found to be -0.700 kcal/mol. Table 7.4 shows the relative conformational

energies of dimethoxymethane, determined by experiment,[8] our forcefield, and MM2(82).[4]

Table 7.4. Relative Conformational Energies of Dimethoxymethane (kcal/mol)

	gauche,gauche	*gauche,trans*	*trans,trans*
Experiment[a]	0	1.71	3.42
This Work	0	1.75	3.31
MM2(82)[b]	0	1.98	4.03

a From reference 8

b From reference 4

The MM2 forcefield, which was not fitted to Uchida and Kubo's experimental data, gives the energy of the *trans,trans* conformer as roughly twice that of the *gauche,trans* - thus supporting the assumption made by Uchida and Kubo in their experiment. It should be noted that although *ab initio* calculations predict the same rank order of conformational energies, they generally give much higher relative energy differences,[7,33,15] with values up to 10.3 kcal/mol obtained for the *trans,trans* form, and 4.55 kcal/mol for the *gauche,trans* form, relative to the *gauche,gauche*.[15]

Because dimethoxymethane possesses only two torsion angles (internal rotational degrees of freedom), the conformational energies may be conveniently expressed as a contour plot. Figure 7.7(a) shows a contour plot of the energy (calculated by our forcefield) as a function of the two **C–O–C–O** torsion angles. The two lowest minima are the *gauche,gauche* conformations at approximately (+60°,+60°) and (-60°,-60°). These are denoted g^+g^+ and g^-g^- in Figure 7.7(a), and are mirror images of each other. The next lowest minima are the *gauche,trans* conformations,

Figure 7.7(a) Calculated Energy Surface for Dimethoxymethane as a Function of Internal Rotation about ϕ and ψ

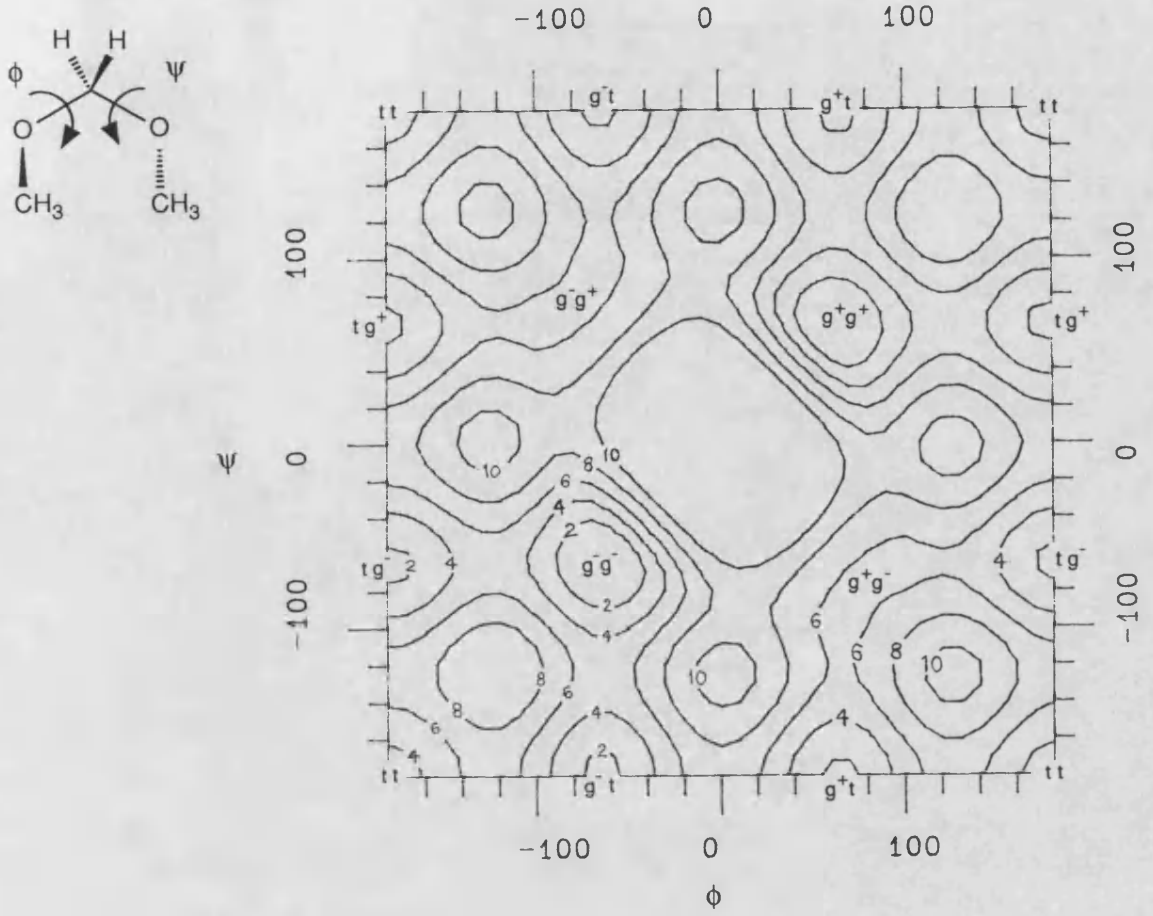
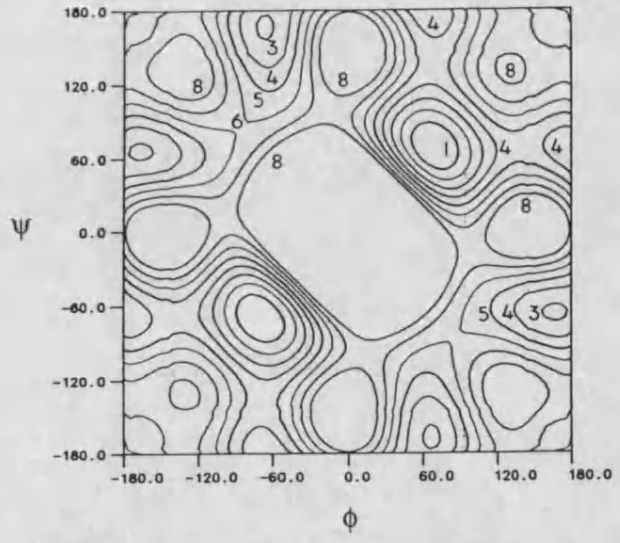
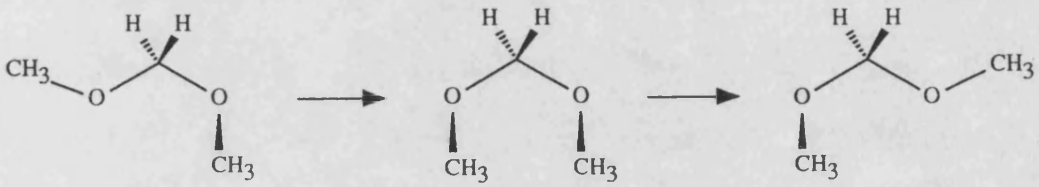


Figure 7.7(b) Energy Surface for Dimethoxymethane calculated by *ab initio* (6-31G*) (reproduced from K.B Wiberg & M.A. Murcko, *J. Am. Chem. Soc.*, **111**, *4821*, (1989))



of which there are four: g^+t, g^-t, tg^+ and tg^-. The *trans,trans* (tt) conformation is the highest energy minimum, and this occurs at the four corners of the plot. The central area of the map is of high energy because this represents geometries were one or both of the torsion angles are close to the eclipsed position. The hypothetical g^+g^- and g^-g^+ conformations near (+60°,-60°) and (-60°,+60°) are not indicated to be minima but rather saddle-points on the transitional pathways between pairs of *gauche-trans* minima.



t g$^+$ g$^-$g$^+$ g$^-$t

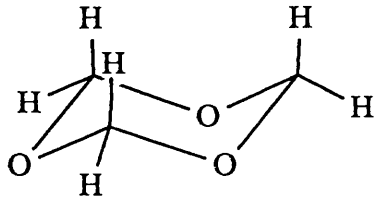
Figure 7.7(b) shows the same energy surface calculated by *ab initio* calculations (6-31G*) and is reproduced here from a recent paper by Wiberg and Murcko.[7] Although the energy values are slightly different, the main features of the two contours plots are very similar.

7.6.2 Vibrational Frequencies of 1,3,5-Trioxane

Although the vibrational spectra of dimethoxymethane have been studied,[34,35] not all of its fundamental frequencies have been observed. This, together with its low symmetry (C_2) and relatively large number of normal modes (33) makes it unsuitable for the fitting of force constant parameters. Instead, the data selected for this purpose were the vibrational frequencies of 1,3,5-trioxane; which can be considered as three

fused acetal units:

C_{3v}



1,3,5-Trioxane

This molecule has the advantage of very high symmetry, giving it a greatly simplified vibrational spectrum. Although trioxane has 12 atoms (giving rise to 30 vibrational modes) symmetry considerations mean that 20 of these modes are in degenerate pairs and therefore give only 10 frequency values. A further 3 frequencies are of the A_2 symmetry species, which are not observed, being inactive in both IR and Raman. These leaves only 17 observed fundamental frequencies in the vibrational spectrum, and of these 4 relate to **C–H** stretching vibrations which are not of interest at present.

The remaining 13 skeletal frequencies were therefore used for determining the eight force constant parameters shown below.

O–C_a	K_b
O–C–O	K_θ, $K_{bb'}$, $K_{b\theta}$
H–(C–O)–O	$K_{\theta\theta'}$
O–(C–H)–O	$K_{\theta\theta'}$
C–O–C–O	V_3, $K_{\theta\theta'\phi}$

On optimisation, the K_b value for the **O–C_a** bond remained close to the initial estimated value (K_b for a standard ether **O–C** bond) and was therefore set to the

same value (104.1 kcal/mol). Another three parameters - the cross term force constants $K_{bb'}$ and $K_{b\theta}$ for the **O–C–O** unit, and $K_{\theta\theta'}$ for the **H–(C–O)–O** angle-angle interaction - were found to have little effect on any of the vibrational frequencies of trioxane, and were therefore set to zero values. The remaining four parameters - **O–C–O**(K_b), **C–O–C–O** (V_3 & $K_{\theta\theta'\phi}$) and **O–(C–H)–O**($K_{\theta\theta'}$) - were assigned values by a least-squares optimisation to the trioxane frequencies. (Parameter values are given in Appendix I.)

Table 7.5 Comparison of Calculated and Experimental Vibrational Frequencies of 1,3,5-Trioxane (cm^{-1}).

Symm.	Calc.	Expt.	Devn.
A_1	2970	2853	117
E	2962	3027	-65
E	2901	2850	51
A_1	2899	2789	110
E	1479	1478	1
A_1	1478	1495	-17
E	1422	1409	13
A_2	1387	-	-
E	1293	1305	-12
A_2	1247	-	-
A_1	1227	1235	-8
E	1172	1174	-2
A_2	1128	-	-
E	1082	1069	13
A_1	964	975	-11
E	922	945	-23
A_1	735	752	-17
E	573	525	48
A_1	470	467	3
E	302	301	1

Experimental data are average values taken from references 36 , 37 , 38 & 39

A comparison between the calculated and experimental frequencies is made in Table 7.5. Several vibrational analyses of trioxane have been made,[36-39] all in generally close agreement, and we have used an *average* of the frequency values from these studies for our comparison. The four **C–H** stretching frequencies show the largest deviations, and this may be due to the anomeric effect causing slight changes in the hybridisation of the carbon atoms. The environment of these **C–H** bonds is therefore different from those in alkanes and simple ethers, and the parameters for these bonds must be considered at the limits of transferability.

The 13 skeletal vibrations (those other than the **C–H** stretches i.e. below 2000 cm^{-1}) were the ones used in the parameter optimisation and are by contrast excellently reproduced. The average deviation for these was only 13.0 cm^{-1}, with a maximum deviation of 48 cm^{-1}.

7.7 Application to Other Acetals

7.7.1 Geometries

Apart from dimethoxymethane, discussed above, there are very few gas phase structural determinations of other acetals. We found only four: 1,3-dioxane, 1,3,5-trioxane, paraldehyde and 2,2-dimethoxypropane. Table 7.6 shows the calculated geometries for these molecules and compares them with experimental values. (The structures of 1,3-Dioxane and Paraldehyde are given in Appendix III.)

1,3-Dioxane. The electron diffraction (r_a) structure for this molecule is poorly determined, having very large experimental uncertainties.[40] Our calculated geometry is generally found to deviate by less than the standard deviation values, which are particularly large for the bond lengths (between 0.009 and 0.028 Å). The electron

Table 7.6 Comparison of Experimental and Calculated Acetal Molecular Geometries[a]

Compound	Internal	Calc.[b,c]	Expt.[c]	Diff.[d]	Reference
1,3-Dioxane	C_2–O	1.409	1.393	0.016	r_a (40)
	C_4–O	1.419	1.439	-0.020	
	C–C	1.523	1.528	-0.005	
	C–O–C	114.6	110.9	3.7	
	O–C–O	112.2	115.0	-2.8	
	C–C–O	112.8	109.2	3.6	
	C–C–C	109.0	107.7	1.3	
	C–C–C–O	±46.8	±57.4	-10.6	
	C–O–C–C	±52.1	±56.0	-3.9	
	C–O–C–O	±54.4	±58.9	-4.5	
1,3,5-Trioxane	C–O	1.404	1.411	-0.007	r_g (41)
	C–H	1.108	1.116	-0.008	
	C–O–C	112.9	109.2	3.7	
	O–C–O	111.7	111.0	0.7	
	C–O–C–O	±52.3	±58.4	-6.1	
Paraldehyde	C–O	1.400	1.410	-0.010	r_a (42)
	C–C	1.522	1.494	0.028	
	C–H	*1.106*	*1.104*	*0.002*	
	C–O–C	113.6	112.3	1.3	
	O–C–O	110.7	110.7	0.0	
	O–C–C	108.5	109.2	-0.7	
	C–O–C–O	±52.8	±54.7	-1.9	
2,2-Dimethoxypropane	C–O	*1.420*	*1.423*	*-0.003*	r_a(43)
	C–C	1.549	1.513	0.036	
	C–O–C	122.1	114.0	8.1	
	O–C–O	112.5	117.4	4.9	
	C–C–C	110.1	112.2	2.1	
	C–O–C–O	54.2	52.0	2.2	
Dimethoxymethane	C–O	*1.412*	*1.407*	*0.005*	r_a(30)
	C–O–C	116.1	114.6	1.5	
	O–C–O	113.9	114.3	-0.4	
	C–O–C–O	64.0	63.3	0.7	

a Bond lengths are in Å, Bond angles in degrees.

b Values in *italics* indicate the internals which are assumed in the experimental model to be equivalent throughout the molecule (e.g. *all* **C–H** lengths in Paraldehyde were assumed to be equal[42]). The appropriate calculated values are averaged to facilitate comparison.

c Experimental values are derived from gas phase electron diffraction data.

d Diff = Calc - Exptl

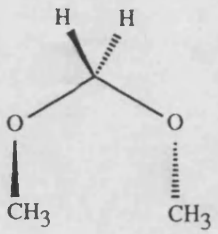
diffraction data for 1,3-dioxane was interpreted as having two very different **C–O** bond lengths (1.393 and 1.439 Å). However, in the optimisation of the geometry to the electron diffraction data, the two **C–O** bond lengths were found to be very highly correlated (0.997) and the exact values of these bond lengths must therefore be in some doubt. A similar problem was discussed for dimethoxymethane (Section 7.6.1) and as in that case, only the average **C–O** bond can be treated as an observable. Taking the average **C–O** bond lengths, the experiment and the calculation are now in good agreement at 1.416 Å and 1.414 Å respectively.

1,3,5-Trioxane. Two gas phase determinations have been made for this molecule, one electron diffraction (r_g)[41] and one microwave.[44] We generally prefer to use electron diffraction results (see Chapter 2) and have chosen the r_g geometry for comparison here. Table 7.6 shows the **C–O** bond lengths to be reasonably well reproduced, as is the rest of the structure with the exception of the **C–O–C** valence angle.

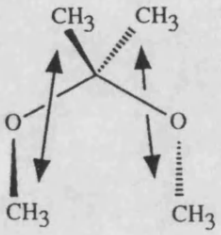
Paraldehyde. This molecule is also known as 2,4,6-trimethyl-trioxane, and is structurally very similar to trioxane. An electron diffraction geometry has been determined,[42] which is found to be in reasonable agreement with our calculated structure in all cases except the **C–C** bond length. The experimentally determined value for this bond length is exceptionally short, even allowing for possible distortions of geometry caused at an anomeric centre. The experimental **C–C** bond length may, in our opinion, be at fault: x-ray data for an analogous compound to paraldehyde, 2,4,6-tri*cyclohexyl*-trioxane shows no such bond shortening; the **C–C** bond length in that case being 1.540 Å.[45]

2,2-Dimethoxypropane. This molecule is structurally similar to dimethoxymethane, but has methyls on the central carbon rather than hydrogens. The average calculated

C–O bond length is close to the experimental value, but generally the remainder of the geometry of this molecule is poorly reproduced. In particular, the **C–O–C** angles were 8.1° larger than the experimentally determined values. (**C–O–C** angles represent a problem for valence forcefields in general, and this was discussed in the previous chapter.) The **C–O–C** angles were opened by strong *gauche* methyl-methyl repulsions occurring in this molecule that are not present in dimethoxymethane.



Dimethoxymethane



2,2-Dimethoxypropane

The MM2(82) forcefield also has problems reproducing the geometry of this molecule, but for this forcefield the problem lies in the **O–C–O** angle, which is overestimated by 7.8°.[4]

7.7.2 Conformational and Configurational Energies

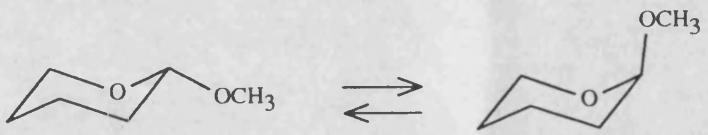
Because of widespread interest in the anomeric effect, many conformational analyses for acetals have been undertaken, both experimentally and theoretically. Most of these studies have tended to focus on cyclic acetals, partly because of the importance of the anomeric effect in pyranose and furanose rings, but also because cyclic systems generally have fewer available conformations and are therefore easier to analyse.

Energy differences are generally determined by population studies of the equilibrium between conformations (or configurations). In most cases, these population studies are carried out in solution; and for acetals in particular, the polarity of solvent

used can have a large effect on the position of the equilibrium and the resulting energy differences.[46,47] As discussed in Chapter 2, for the purposes of comparison with our calculated energy differences, experimental data should ideally relate to energy differences in dilute, non-polar solutions, where solvent effects will be minimised.

The calculated and experimental energy differences for a range of acetals is discussed below.

2-Methoxytetrahydropyran. The axial-equatorial energy difference for this molecule has received a great deal of attention because it is the most basic model for the study of the anomeric effect in pyranose rings. Early results indicated the axial-equatorial energy difference in non-polar solvents to be about -1 kcal/mol (i.e in favour of the axial conformer).[48,46]



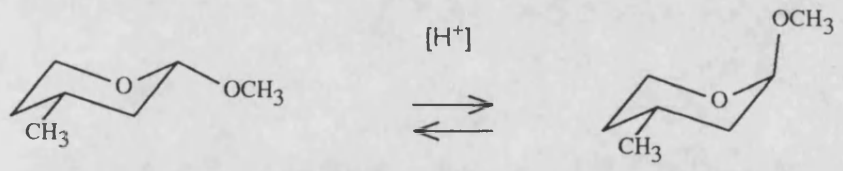
equatorial axial

In more recent experiments by Booth,[49-51] however, results seemed to indicate that the enthalpy difference between the two conformers was about zero, while the preference for the axial conformer was due to entropic effects. Booth carried out his experiments using variable temperature NMR with a $CFCl_3/CDCl_3$ (85:15) solvent mixture; and in the light of further work by Lemieux, it seems likely that this choice of solvent may be partly responsible for these results. Prompted by Booth's results, Lemieux studied the effects of different solvents on the axial/equatorial equilibrium for

2-methoxytetrahydropyran, and found that the preference for the axial form was far less pronounced in polar solvents, especially those that had polar hydrogens (or deuteriums, like $CDCl_3$) that could form hydrogen bonds.[52] In non-polar solvents (CCl_4/C_6D_6) Lemieux obtained a $\Delta H_{eq \rightarrow ax}$ value of -0.8 kcal/mol (in closer agreement with the early experiments[48]) and estimated the enthalpy difference for the isolated molecule to be about -1 kcal/mol.

Using our forcefield, we calculate an axial-equatorial energy difference of -1.06 kcal/mol, in excellent agreement with both Lemieux's results and the earlier ones of de Hoog *et al.*[48] Other computational methods also give similar values; *ab initio* calculations using the 6-31G* basis set give an energy difference of -1.33 kcal/mol,[7] while the MM2(82) forcefield gives a value of -1.17 kcal/mol.[4]

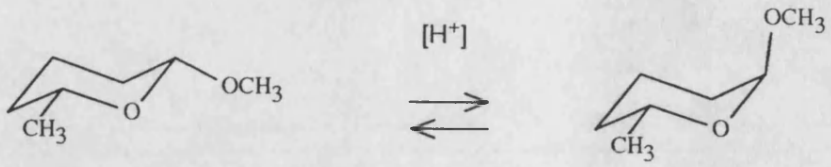
cis- and trans- 2-Methoxy-4-methyltetrahydropyran. The relative energies of these two configurations were studied by equilibration in CCl_4 in the presence of mineral acid.[47] The relative proportions of each configuration was then established by gas chromatography of the mixture, and $\Delta G_{eq \rightarrow ax}$ found to be -0.83 kcal/mol. This is in reasonable agreement with the calculated $\Delta E_{eq \rightarrow ax}$ value of -1.42 kcal/mol.



equatorial axial

cis- and trans- 2-Methoxy-6-methyltetrahydropyran. The $\Delta G_{eq \rightarrow ax}$ value for this equi-

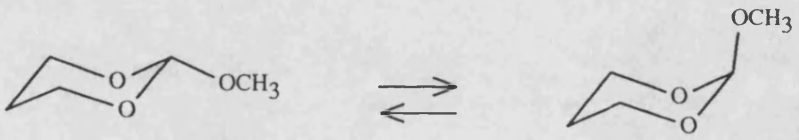
librium was studied by the same method as the 4-methyl analogue (above).[47]



equatorial axial

The experimental ($\Delta G_{eq \rightarrow ax}$) and calculated ($\Delta E_{eq \rightarrow ax}$) values are similar to those for the 4-methyl analogue, at -0.73 kcal/mol and -1.44 kcal/mol respectively.

2-Methoxy-1,3-dioxane. This molecule was determined from dipole moment measurements in benzene to favour the axial conformation ($\Delta G_{eq \rightarrow ax}$ = -0.62 kcal/mol).[53] This is to be expected since it experiences a 'double' anomeric effect as there are two ring oxygens present.

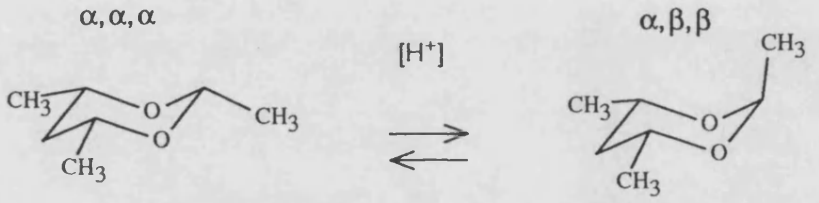


equatorial axial

The calculated energy difference (-0.85 kcal/mol) is in good agreement with the experimental value.

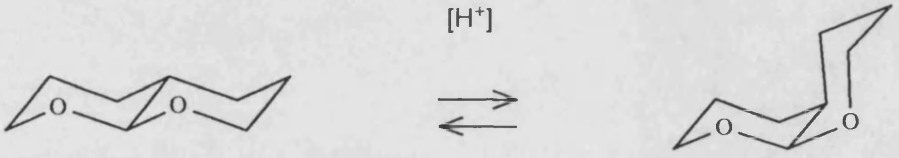
α,β,β- *and* α,α,α- *2,4,6-Trimethyl-1,3-dioxane.* The two configurations of this

equilibrium are shown below. Because there is no anomeric effect for these molecules, the equilibrium is governed by normal steric effects, and the more stable configuration is the equatorial (α,α,α) form. The experimental value for $\Delta G_{eq \rightarrow ax}$ (determined in diethylether at 25°C) is +3.98 kcal/mol.[53] The calculation also shows the equatorial form to be the more stable, but the energy difference is somewhat smaller ($\Delta E_{eq \rightarrow ax}$ +2.51 kcal/mol).



equatorial axial

cis- and trans- 1,8-Dioxadecalin. Beaulieu *et al.* studied the equilibrium between these two configurations in methanol, and found the *cis:trans* ratio at room temperature to be 55:45.[54] Allowing for the two equivalent conformations of the *cis* configuration (ΔS=RT ln2) the $\Delta H_{trans \rightarrow cis}$ value is estimated to be 0.30 kcal/mol.



trans-1,8-dioxadecalin cis-1,8-dioxadecalin

Contrary to experiment, however, the calculations for these molecules indicate the *cis* form to be the more stable (by 1.25 kcal/mol). This discrepancy may be the result of

solvent effects occurring in the experiment. The *trans* form is calculated to have a a higher dipole moment than the *cis* (3.8 D *versus* 2.8 D) and may therefore be stabilised more in methanol. This argument concurs with the results of Lemieux, which showed that polar solvents reduced the preference for the axial form of 2-methoxytetrahydropyran.[52]

Hydrogen bonding may also play a part in this equilibrium. The folded shape of the *cis* form may restrict the number of methanol molecules that could solvate the oxygen atoms. Since hydrogen bond energies are in the range 4-5 kcal/mol, effects like this would have a marked influence on energy differences of this size.

7.8 References to Chapter 7

1. R.U. Lemieux, in *Molecular Rearrangements, Part 2*, ed. P. de Mayo, Interscience, New York (1964).
2. A.J. Kirby, in *The Anomeric Effect and Related Stereoelectronic Effects at Oxygen*, Springer, Berlin (1983).
3. S. Wolfe, M-H. Whangbo, and D.J. Mitchell, *Carbohydr. Res.*, **69**, 1 (1979).
4. L. Norskov-Lauritsen and N.L. Allinger, *J. Comput. Chem.*, **5**, 326 (1984).
5. I. Tvaroska and S. Perez, *Carbohydr. Res.*, **149**, 389 (1986).
6. U. Burkert, *Tetrahedron*, **35**, 1945 (1979).
7. K.B. Wiberg and M.A. Murcko, *J. Am. Chem. Soc.*, **111**, 4821 (1989).
8. T. Uchida, Y. Kurita, and M. Kubo, *J. Polym. Sci.*, **19**, 365 (1956).
9. J. R. Durig and D. A. C. Compton, *J. Chem. Phys.*, **69**, 4713-4719 (1978).
10. L.O. Brockway, *J. Phys. Chem.*, **41**, 185 (1937).
11. G.A. Jeffrey , J.A. Pople, and L. Radom, *Carbohydr. Res.*, **25**, 117 (1972).

12. H.M. Berman, S.S.C. Chu, and G.A. Jeffrey, *Science*, **157**, 1576 (1967).

13. C. Romers, C. Altona, H.R. Buys, and E. Havinga, *Topics in Stereochemistry*, **4**, 39, Wiley-Interscience (1969).

14. G.A. Jeffrey, J.A. Pople, and L. Radom, *Carbohydr. Res.*, **38**, 81 (1974).

15. C. Van Alsenoy, L. Schafer, J.N. Scarsdale, and J.O. Williams, *J. Mol. Struc. THEOCHEM*, **86**, 111 (1981).

16. J.T. Edward, *Chemistry & Industry*, 1102 (1955).

17. R.O. Hutchins, L.D. Kopp, and E.L. Eliel, *J. Am. Chem. Soc.*, **90**, 7174 (1968).

18. E.A.C. Lucken, *J. Chem. Soc., III*, 2954 (1959).

19. L. Radom, W.J. Hehre, and J.A. Pople, *J. Am. Chem. Soc.*, **94**, 2371 (1972).

20. G.A. Jeffrey and J.H. Yates, *J. Am. Chem. Soc.*, **101**, 820 (1979).

21. P. Deslongchamps, in *Stereoelectronic Effects in Organic Chemistry*, Pergamon, Oxford (1983).

22. A. Cosse-Barbi, D.G. Watson, and J.E. Dubois, *Tetrahedron Lett.*, **30**, 163 (1989).

23. D.A. Sweigart and D.W. Turner, *J. Am. Chem. Soc.*, **94**, 5599 (1972).

24. T. Koybayashi and S. Nagakuru, *Bull. Chem. Soc. Jpn.*, **46**, 1558 (1973).

25. R.U. Lemieux and S. Koto, *Tetrahedron*, **30**, 1933 (1974).

26. H. Thogersen, R.U. Lemieux, K. Bock, and B. Meyer, *Can. J. Chem.*, **60**, 44 (1982).

27. R.U. Lemieux and K. Bock, *Archiv. Biochem. Biophys.*, **221**, 125 (1983).

28. N.L. Allinger and D.Y. Chung, *J. Am. Chem. Soc.*, **98**, 6298 (1976).

29. Figure 7.3 shows the shape of the surface at values of b relatively close to b_0. At larger deviations from b_0, the bond stretch energy is not symmetrical about b_0 (see Figure 3.1).

30. E.E. Astrup, *Acta Chem. Scand.*, **27**, 3271 (1973).

31. S. R. Niketic and K. Rasmussen, in *The Consistent Force Field*, Springer, New York (1977).

32. L. Schafer, C. Van Alsenoy, and J.N. Scarsdale, *J. Mol. Struc. THEOCHEM*, **86**, 349 (1982).

33. G.A. Jeffrey, J.A. Pople, J.S. Binkley, and S. Vishveshwara, *J. Am. Chem. Soc.*, **100**, 373 (1978).

34. K. Nukada, *Spectrochim. Acta*, **18**, 745 (1962).

35. J.K. Wilmshurst, *Can. J. Chem.*, **36**, 285 (1958).

36. W.R. Ward, *Spectrochim. Acta*, **21**, 1311 (1965).

37. A.T. Stair and J.R. Nielsen, *J. Chem. Phys.*, **27**, 402 (1957).

38. M. Kobayashi, R. Iwamoto, and H. Tadokoro, *J. Chem. Phys.*, **44**, 922 (1966).

39. H.M. Pickett and H.L. Strauss, *J. Chem. Phys.*, **53**, 376 (1970).

40. G. Schultz and I. Hargittai, *Acta Chim. Acad. Scient. Hung.*, **83**, 331 (1974).

41. A.H. Clark and T.G. Hewitt, *J. Mol. Struc.*, **9**, 33 (1971).

42. E.E. Astrup, *Acta Chem. Scand.*, **27**, 1345 (1973).

43. E.E. Astrup and A.M. Aomar, *Acta Chem. Scand. A*, **29**, 794 (1975).

44. J.M. Colmont, *J. Mol. Struc.*, **21**, 387 (1974).

45. G. Diana and P. Ganis, *Atti. Acad. Nazion. Lincei*, **35**, 80 (1963).

46. R.U. Lemieux, A.A. Pavia, J.C. Martin, and K.A. Watanabe, *Can. J. Chem.*, **47**, 4425 (1969).

47. E.L. Eliel and C.A. Giza, *J. Org. Chem.*, **33**, 3754 (1968).

48. A.J. de Hoog, H.R. Buys, C. Altona, and E. Havinga, *Tetrahedron*, **25**, 3365 (1969).

49. H. Booth, T.B. Grindley, and K.A. Khedhair, *J. Chem. Soc., Chem. Commun.*, 1047 (1982).

50. H. Booth and K.A. Khedhair, *J. Chem. Soc., Chem. Commun.*, 467 (1985).

51. H. Booth, K.A. Khedhair, and S.A. Readshaw, *Tetrahedron*, **43**, 4699 (1987).

52. J.-P. Praly and R.U. Lemieux, *Can. J. Chem.*, **65**, 213 (1987).

53. F.W. Nader and E.L. Eliel, *J. Am. Chem. Soc.*, **92**, 3050 (1970).

54. N. Beaulieu, R.A. Dickinson, and P. Deslongchamps, *Can. J. Chem.*, **58**, 2531 (1980).

Concluding Remarks

The Forcefield

The forcefield described in this thesis has in most respects been demonstrated to fulfil the original objectives: that is, to reproduce the structural and energetic aspects of molecular behaviour for a range of carbohydrate model compounds.

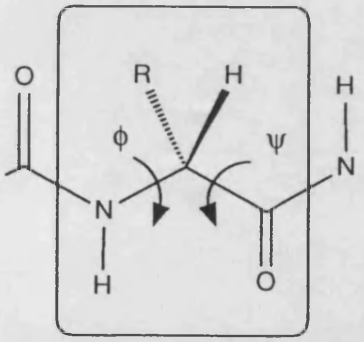
Problems encountered in deriving the forcefield originated mainly from the 1,4 interactions. These are the interactions between atoms separated by three bonds. In hydrocarbons, 1,4 interactions could be represented sufficiently well by considering the non-bond terms between atoms 1 and 4, and a threefold torsional term. However, the introduction of oxygen atoms has meant that additional terms, in particular a one-fold torsional term, are necessary to reproduce the orbital effects caused by the presence of the oxygen lone-pairs. An extreme example of these orbital effects is of course the anomeric effect, which also required the introduction of bond-torsion cross terms (Chapter 7). We foresee that 1,4 interactions will be a recurring problem when forcefields are being developed for new classes of compounds, since they represent something of a 'halfway house' between pure non-bond and bonded interactions.

In alcohols, difficulties with 1,4 interactions were compounded by the neglect of van der Waals effects for hydroxyl hydrogens. While this gives reasonable results for hydrogen bonded crystals (Chapter 5), it probably contributed to the necessity for additional parameters to reproduce the rotameric energies and **C–O** bond lengths in alcohols (Chapter 6). Because of the importance of hydrogen bonding in carbohydrate conformation, the treatment of the non-bond interactions of hydroxyl hydrogens may benefit from further study.

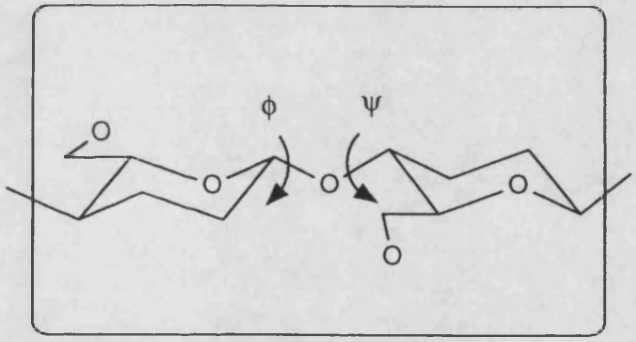
Future Applications

The work presented in this thesis covers only the derivation of a carbohydrate forcefield. Although all the parameters necessary for modelling standard carbohydrates have been developed, time has not permitted the application of the forcefield to a real carbohydrate system.

There are many possible applications for a carbohydrate forcefield. One in particular is the derivation of 'Ramachandran Maps' for disaccharide linkages, similar to those often used to rationalise peptide conformation.[1] An amino acid residue in a peptide chain has basically two conformational degrees of freedom - internal rotation about ϕ and ψ (since the amide bond, ω, is fixed due to its double bond character). The conformational energy of an amino acid residue can therefore be conveniently displayed as a contour map with axes ϕ and ψ. (cf. Figure 7.7(a), p.170).



amino acid residue



disaccharide linkage

Similar maps may be constructed to help in understanding polysaccharide conformation. Flexibility in polysaccharides stems from the glycosidic linkages, since the pyranose rings are relatively rigid 'chairs'. Each glycosidic linkage has two conformational degrees of freedom, ϕ and ψ, and linkages between different pyranose residues will therefore show preferences for different ϕ/ψ values. Ramachandran maps may therefore be calculated using the new forcefield for commonly occurring disaccharide linkages.

Another interesting application of the forcefield would be in the study of a series of oligosaccharide plant hormones (the 'oligosaccharins') identified by Albersheim and Darvill.[2] A range of heptaglucosides (saccharides consisting of seven glucose residues) were found to have different activities depending on how the saccharide chain was branched. The new forcefield could be used, in conjunction with molecular dynamics, to study the likely conformations of these heptaglucosides in order to throw some light on their structure-activity relationships.

References

1. G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, *J. Mol. Biol.*, **7**, 95 (1963).

2. P. Albersheim and A.G. Darvill, *Scient. Am.*, 44 (Sept. 1985).

Appendix I

Forcefield Parameters

Atom Types

Atom Type	Description	Default Type	Rel. Mass
C	General tetravalent carbon atom	None	12
C_a	Anomeric carbon atom (tetravalent carbon atom bonded to two or more oxygens)	**C**	12
C_6	Tetravalent carbon atom in a six-membered ring	**C**	12
O	Ether oxygen atom	None	16
O_H	Hydroxyl oxygen atom	**O**	16
H	Aliphatic hydrogen atom	None	1
H_O	Hydroxyl hydrogen atom	**H**	1

The above table gives the seven atom types used in the forcefield. The following tables give the forcefield parameters; if no parameter is specified for a particular atom type, then the equivalent parameter for the *default* atom type (above) is used. (For example, there are no specific parameter values for a **C_a–H** bond, and so the parameters for the **C–H** bond are used.)

Bond Parameters

Bond	K_b	b_0	α
C–H	108.6	1.105	1.771
C–C	88.0	1.528	1.915
C_6–C_6	88.0	1.504	1.915
C_a–C_6	88.0	1.504	1.915
C–O	104.1	1.409	1.915
C_6–O	104.1	1.390	1.915
C–O_H	104.1	1.424	1.915
C_a–O	104.1	1.381	1.915
O_H–H_O	104.0	0.943	2.280

Angle Parameters

Angle	$\frac{1}{2}K_\theta$	θ_0	$K_{bb'}$	$K_{b\theta}$	$K_{b'\theta}$
H–C–H	39.5	106.4	0.0	0.0	0.0
H–C–C	44.0	110.0	0.0	0.0	25.0
C–C–C	31.8	113.0	16.25	0.0	0.0
H–C–O	58.4	110.0	0.0	0.0	27.7
O–C–O	72.2	108.5	0.0	0.0	0.0
C–C–O	56.3	113.0	16.25	0.0	0.0
C–O–C	50.2	104.0	16.25	0.0	0.0
C–O–H	55.1	106.4	0.0	32.0	0.0

For the angle **A–B–C**; b represents the length of **A–B**, and b′ the length of **B–C**

Torsion Parameters

Torsion	V_1	V_2	V_3	$K_{\theta\theta'\phi}$	$K_{b\phi}$	$K_{b'\phi}$
C–C–C–C	0.0	0.0	0.304	0.0	0.0	0.0
C–C–C–H	0.0	0.0	0.126	-8.1	0.0	0.0
C–C–C–O	-0.89	0.0	0.304	0.0	0.0	0.0
H–C–C–H	0.0	0.0	0.134	-13.4	0.0	0.0
H–C–C–O	0.0	0.0	0.247	-17.5	0.0	0.0
O–C–C–O	-3.55	0.0	1.069	0.0	0.0	0.0
H–C–O–C	0.0	0.0	0.218	-21.1	0.0	0.0
C–C–O–C	0.0	0.0	0.304	-9.1	0.0	0.0
O–C–O–C	-0.70	0.0	1.360	-15.0	3.76	6.97
O–C–O–H	-0.70	0.0	1.360	-15.0	3.76	6.97
H–C–O–H	0.0	0.0	0.165	-14.6	0.0	0.0
C–C–O–H	1.0	0.0	0.304	0.0	0.0	0.0

For the torsion **A–B–C–D**; b represents the length of **B–C**, and b′ the length of **A–B**

Angle-Angle Cross Term Parameters

Angle-Angle[a]	$K_{\theta\theta'}$
C–(C–C)–C	-7.9
H–(C–H)–H	0.0
H–(C–H)–C	0.0
H–(C–C)–H	-7.9
H–(C–C)–C	-7.9
C–(C–H)–C	0.0
H–(C–O)–H	-4.6
H–(C–H)–O	4.5
H–(C–C)–O	-7.9
H–(C–O)–C	-7.9
C–(C–H)–O	11.4
C–(C–C)–O	-7.9
C–(C–O)–C	-7.9
O–(C–H)–O	19.0
H–(C–O)–O	0.0
O–(C–C)–O	-7.9
C–(C–O)–O	-7.9
O–(C–O)–O	-7.9

a The notation **A–(B–C)–D** represents the cross term between angles **A–B–C** and **C–B–D**.

Non-Bond Parameters

Atom Type	r^*_{ii}	ε_{ii}	q_i
C	4.35	0.039	*a*
O	3.21	0.228	-0.38
H	2.75	0.038	+0.10
H_O	0	0	+0.35

a Partial charges on **C** atoms are calculated from: $q_C = 0.19n_O + 0.03n_{OH} - 0.10n_H$ (see Chapter 5)

Parameter Units

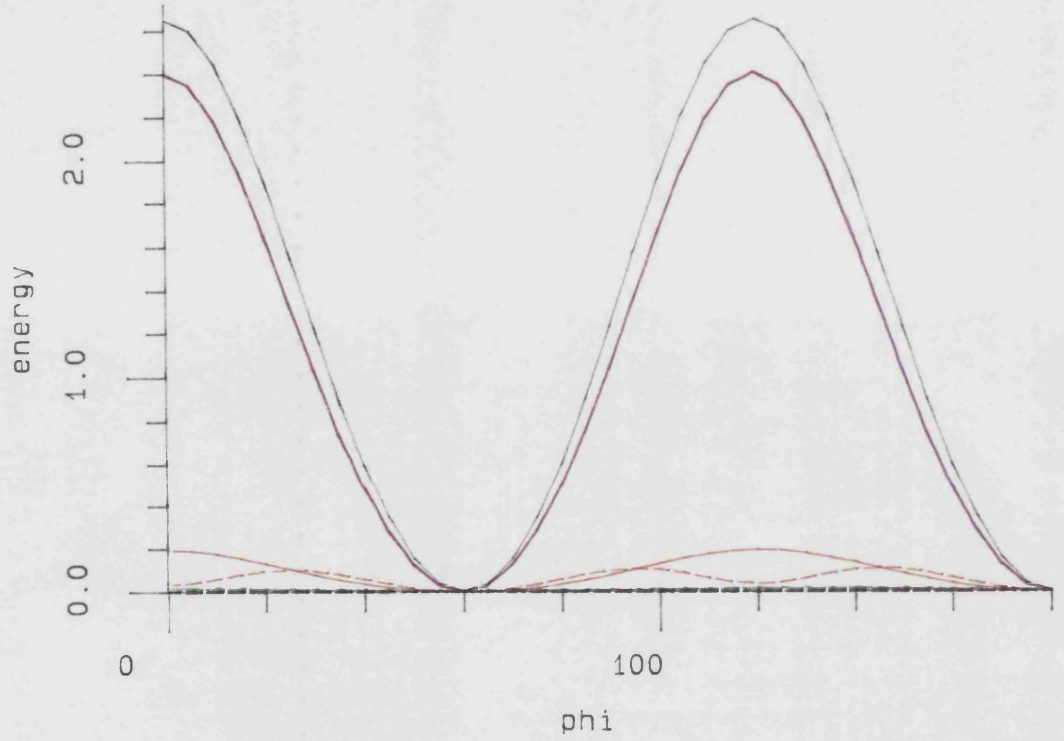
Parameters	Units
b_0, r_{ii}^*	Å
θ_0	degrees
K_b, V_1, V_2, V_3, ε	kcal/mol
α	$Å^{-1}$
K_θ, $K_{\theta\theta'\phi}$, $K_{\theta\theta'}$	kcal/mol/rad^2
$K_{bb'}$	kcal/mol/Å^2
$K_{b\theta}$, $K_{b'\theta}$	kcal/mol/Å/rad
$K_{b\phi}$, $K_{b'\phi}$	kcal/mol/Å
q_i	electronic charge (*e*)

Appendix II

Rotational Barrier Plots

Figure 1

Rotational Barrier for Ethane C-C bond



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 2

Rotational Barrier for n-Butane (C-C-C-C)

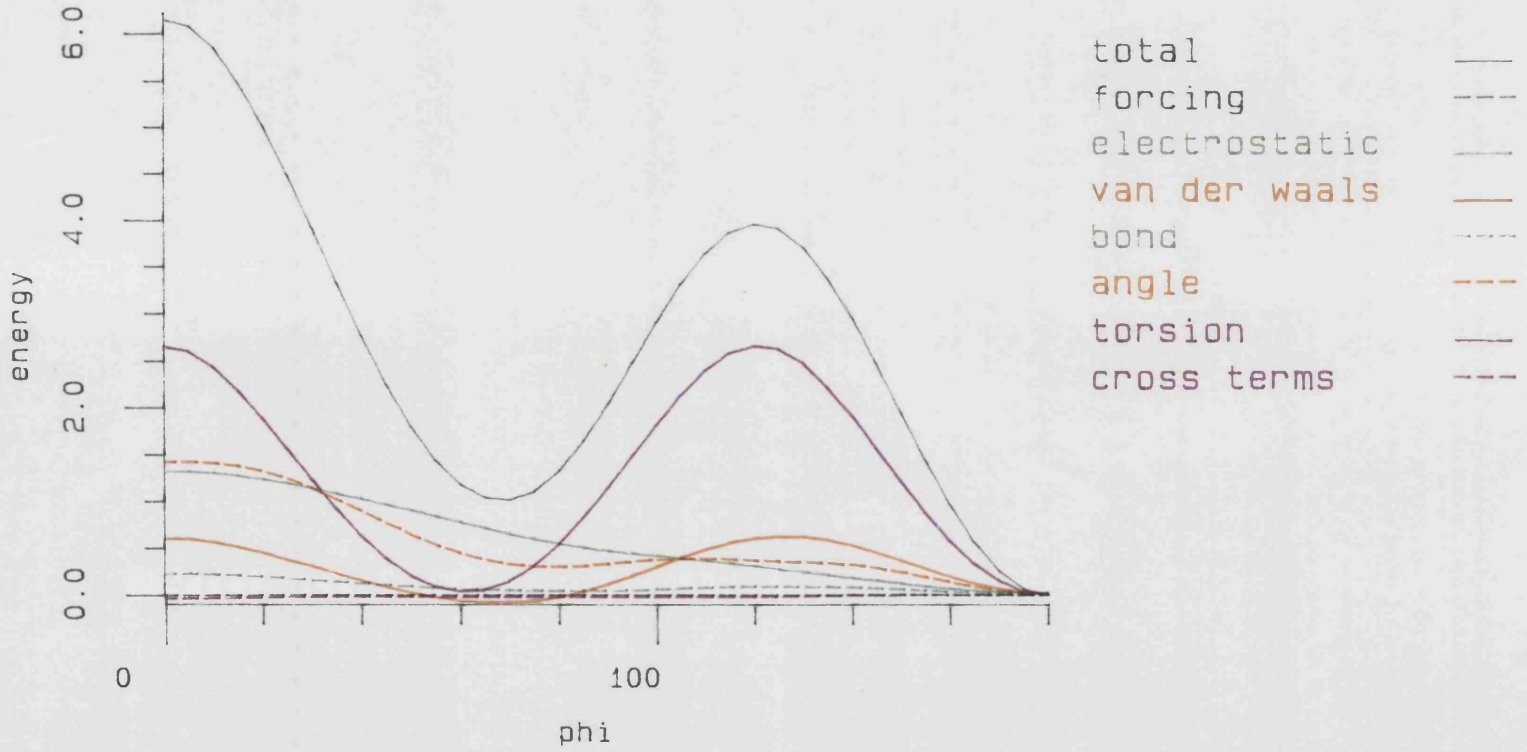


Figure 3

Rotational Barrier for 2-Methylbutane (H-C2-C3-C4)

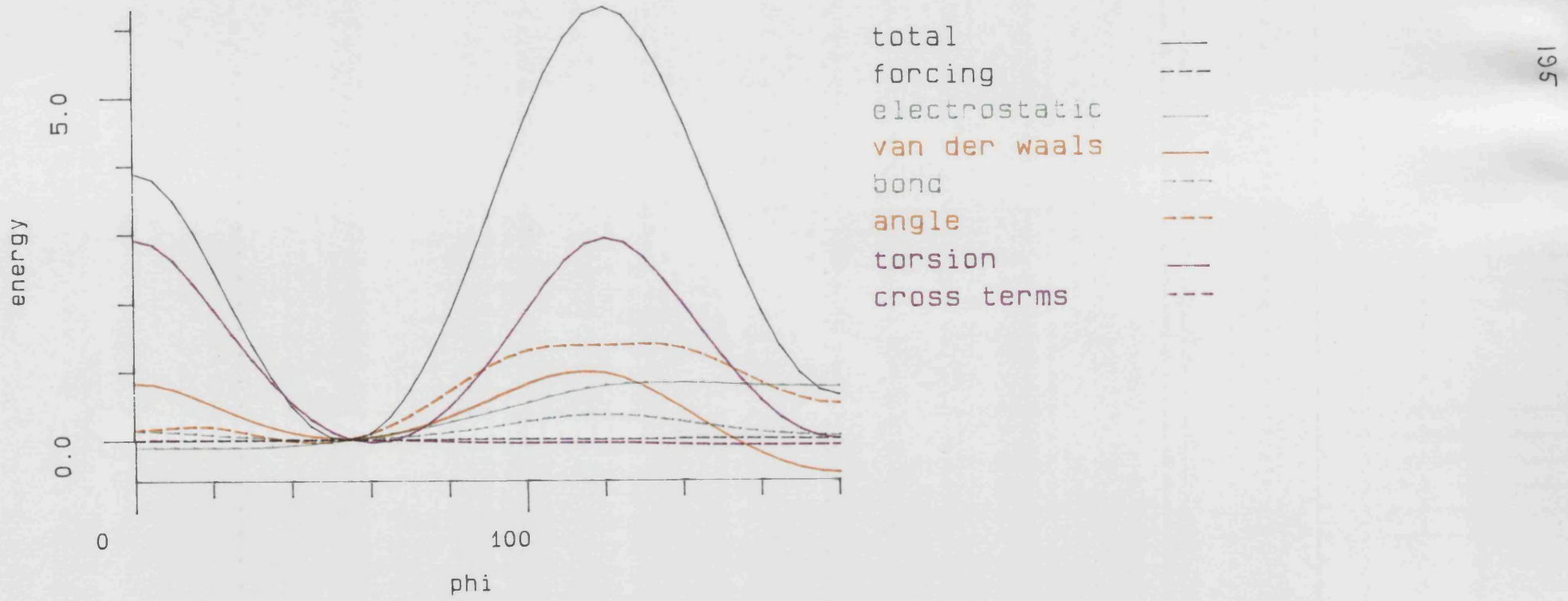
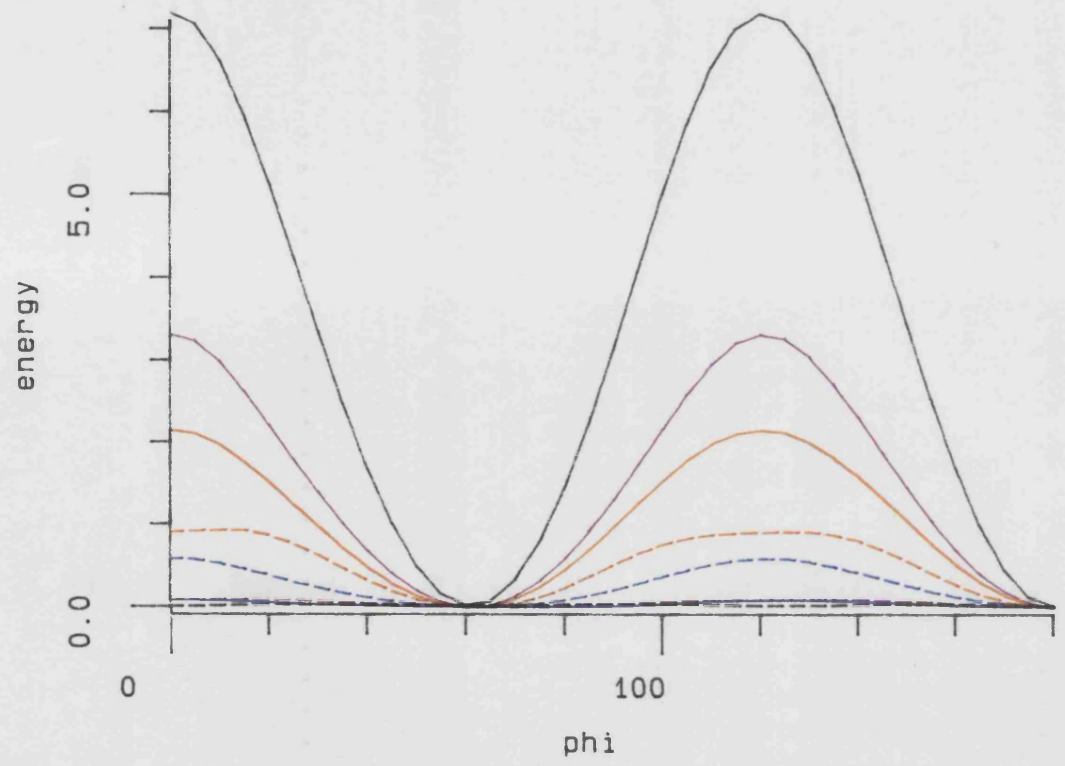


Figure4

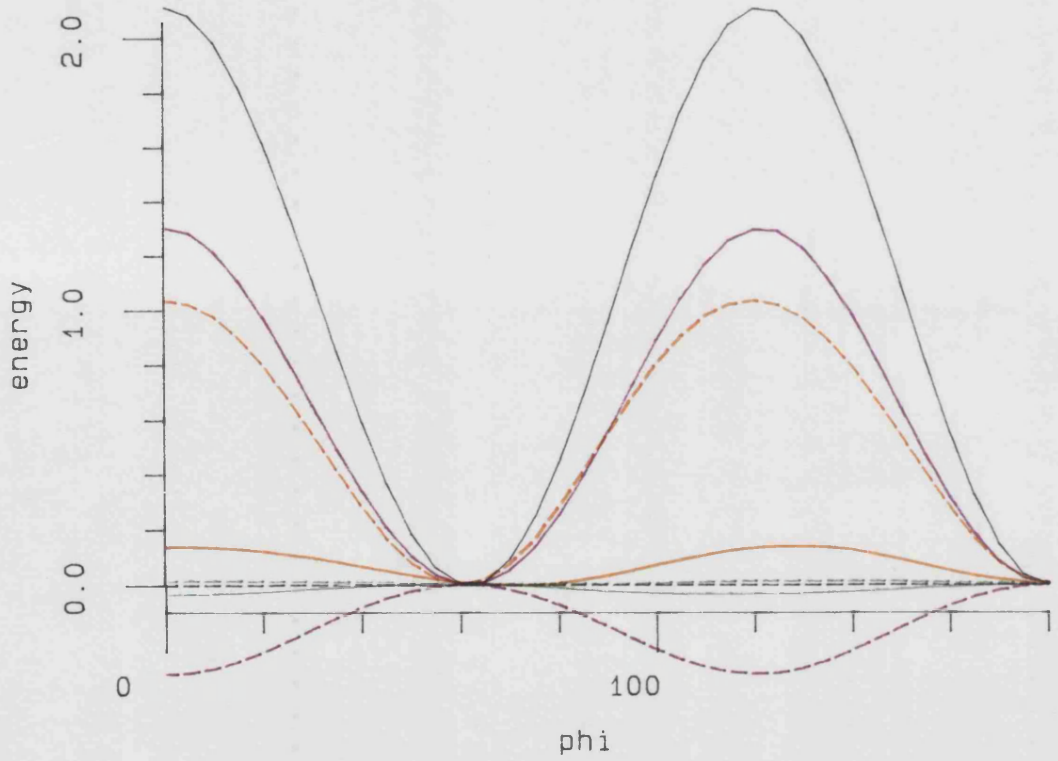
Rotational Barrier for 2,2-Dimethylbutane (C1-C2-C3-C4)



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 5

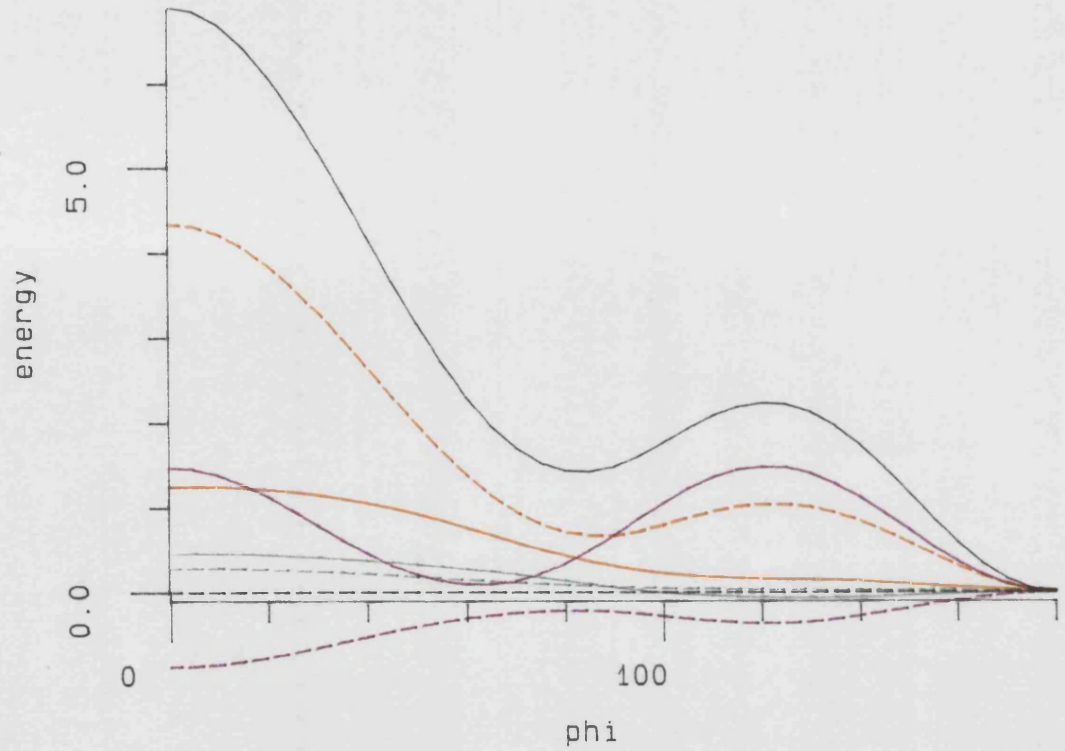
Rotational Barrier for Dimethylether (H-C-O-C)



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 6

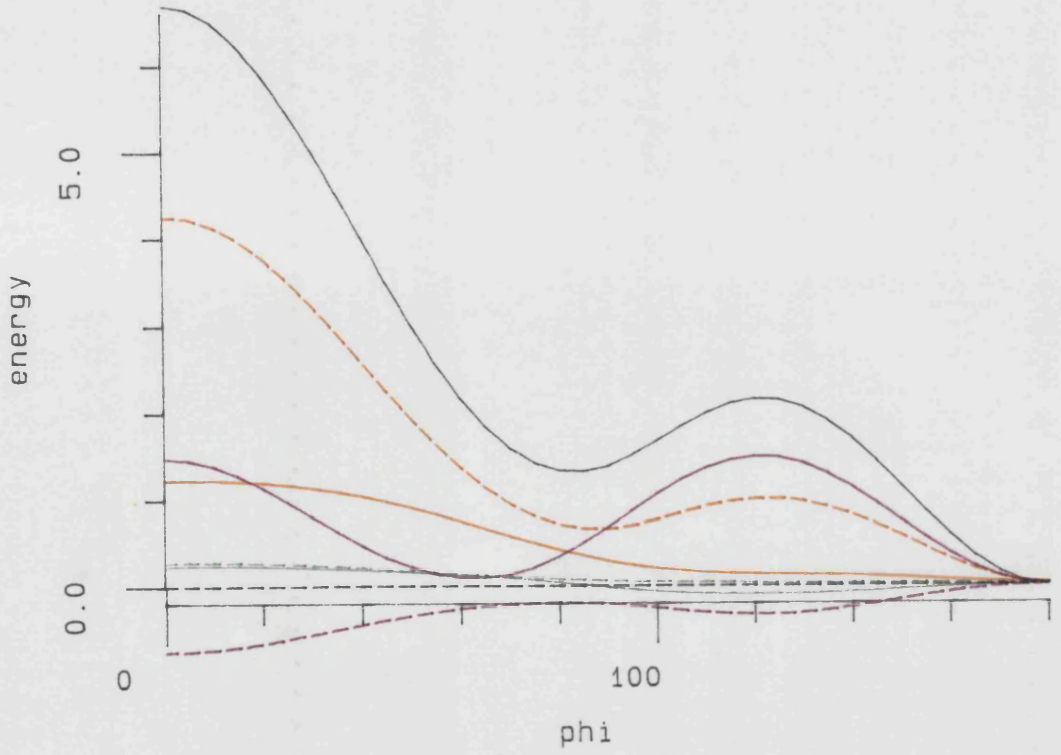
Rotational Barrier for Ethylmethylether (C-O-C-C)



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 7

Rotational Barrier for Diethylether (C-C-O-C)



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 8

Rotational Barrier for i-Propylmethylether (H-C-O-CH3)

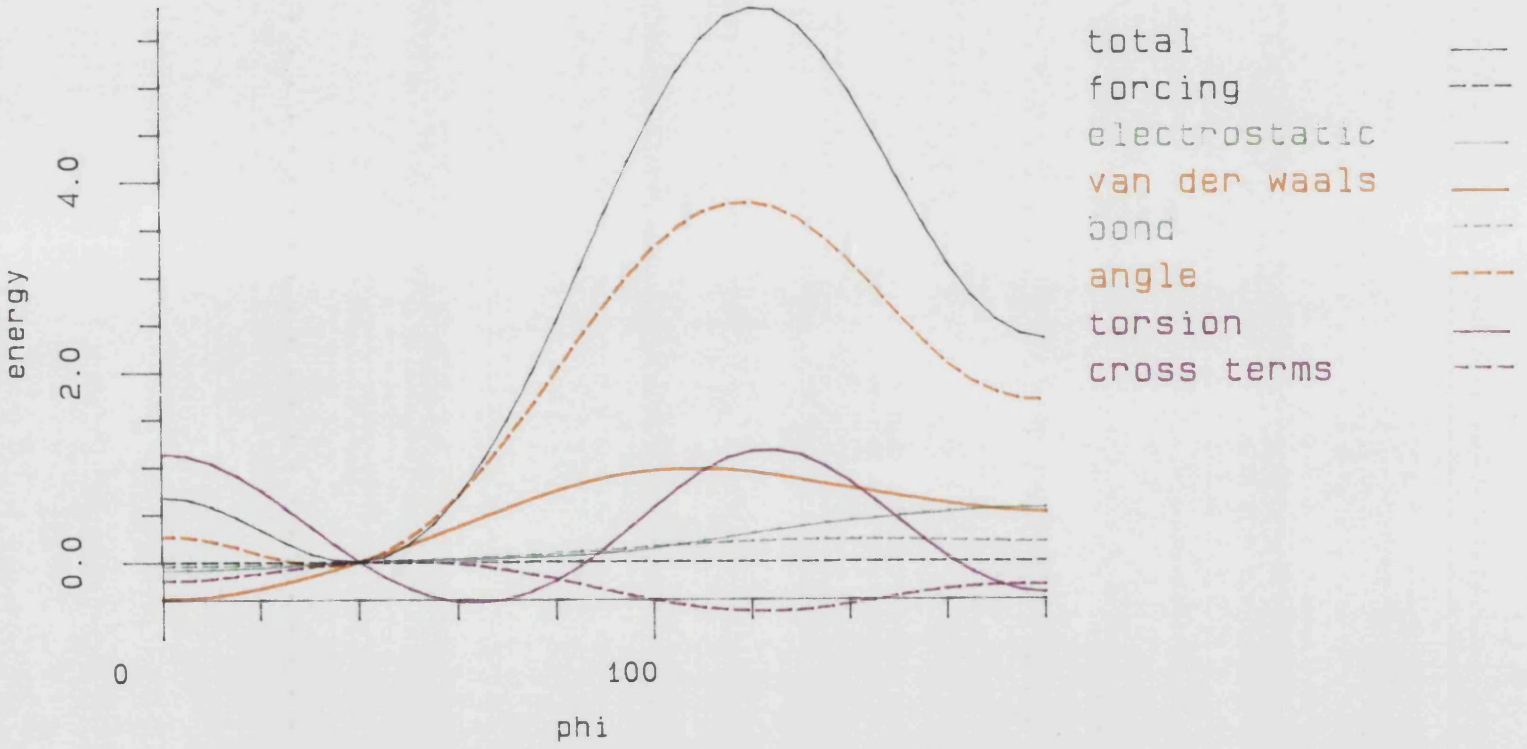


Figure 9

Rotational Barrier for t-Butylmethylether (H-C-C-O)

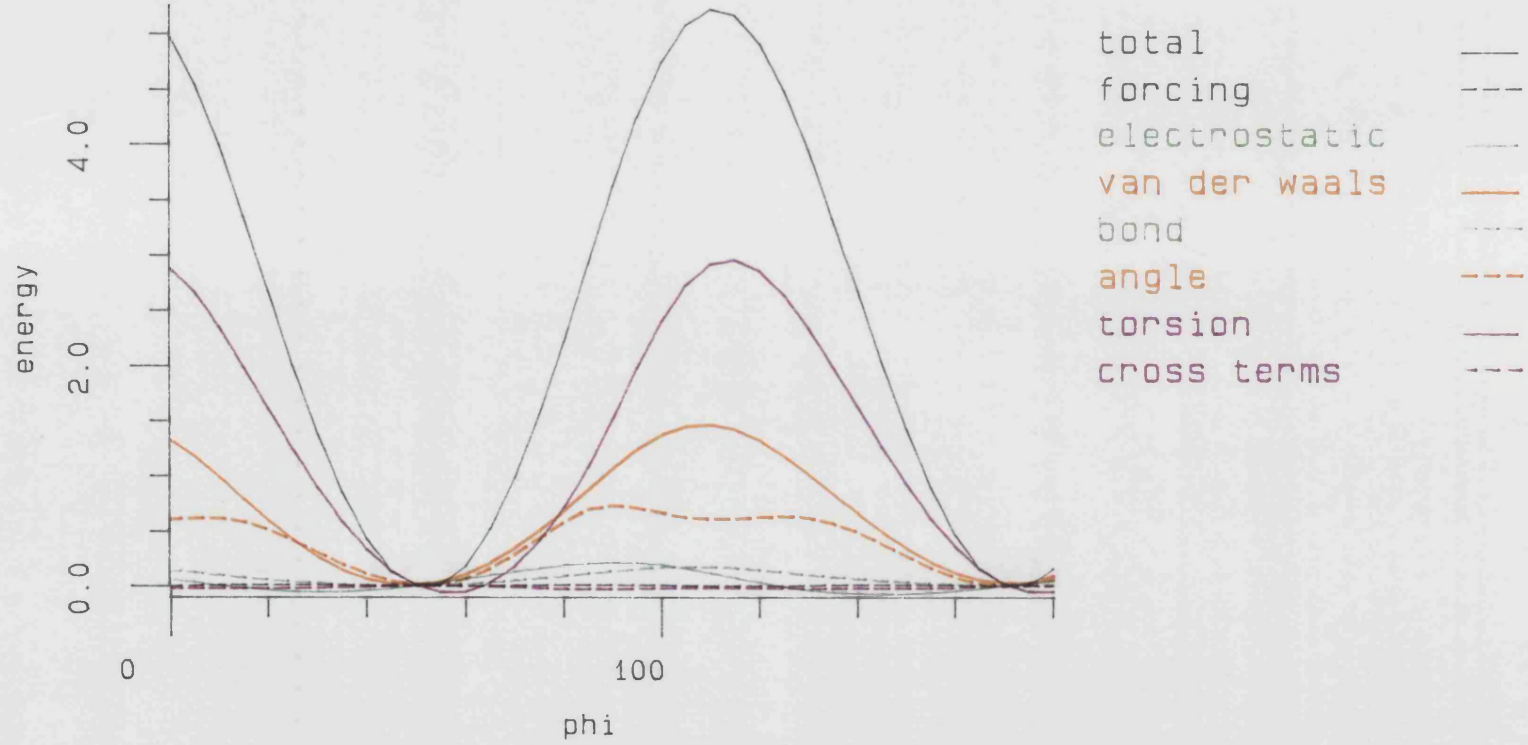
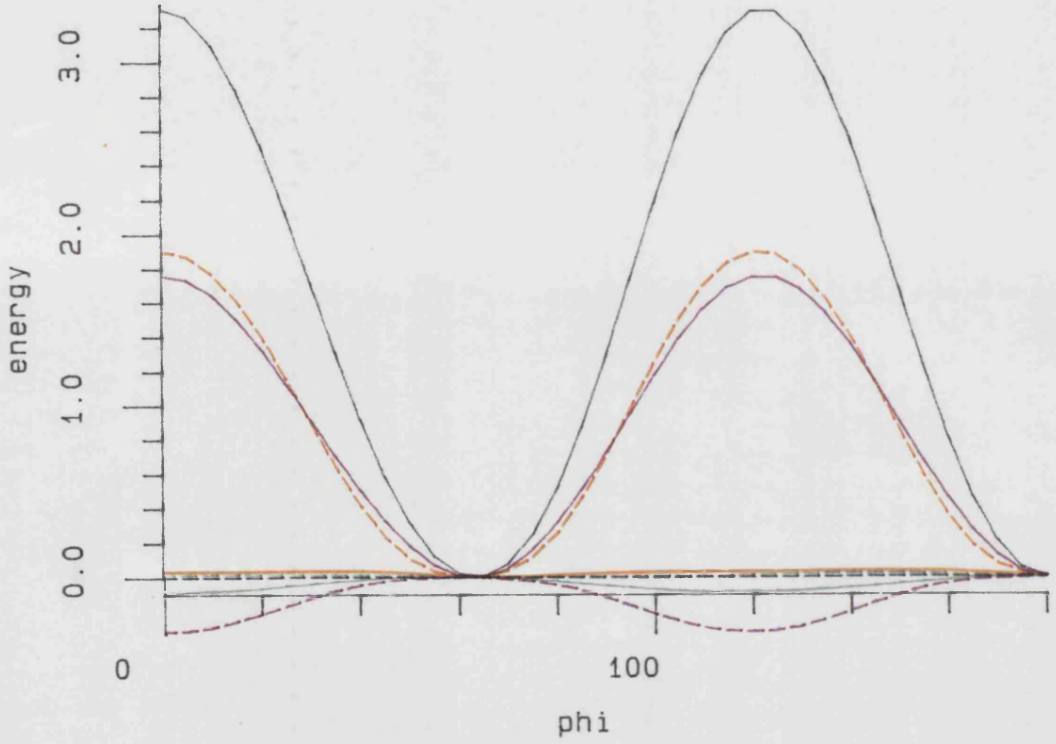


Figure 10

Rotational Barrier for t-Butylmethylether (C-C-O-C)



total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 11

Rotational Barrier for 1,2-Dimethoxyethane (O-C-C-O)

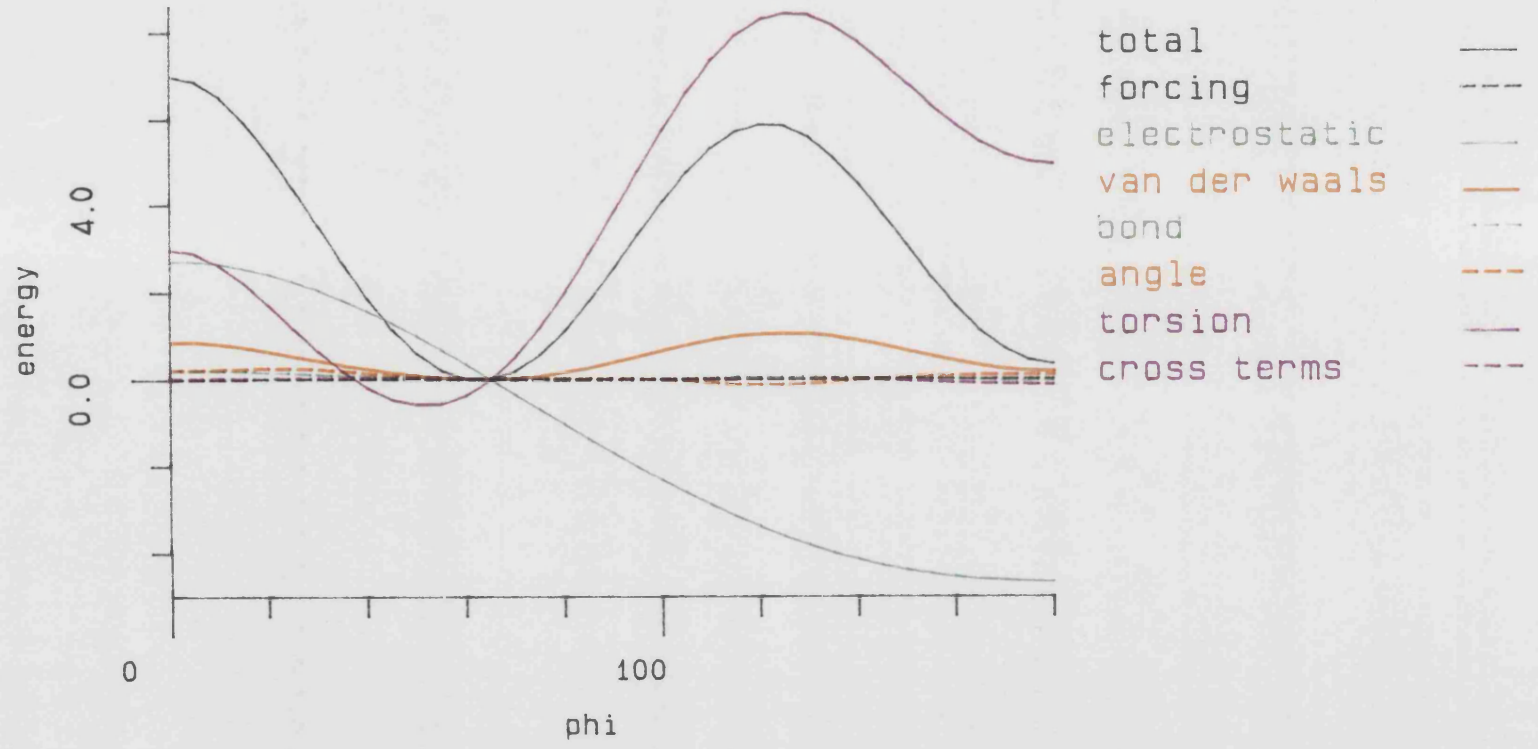


Figure 12

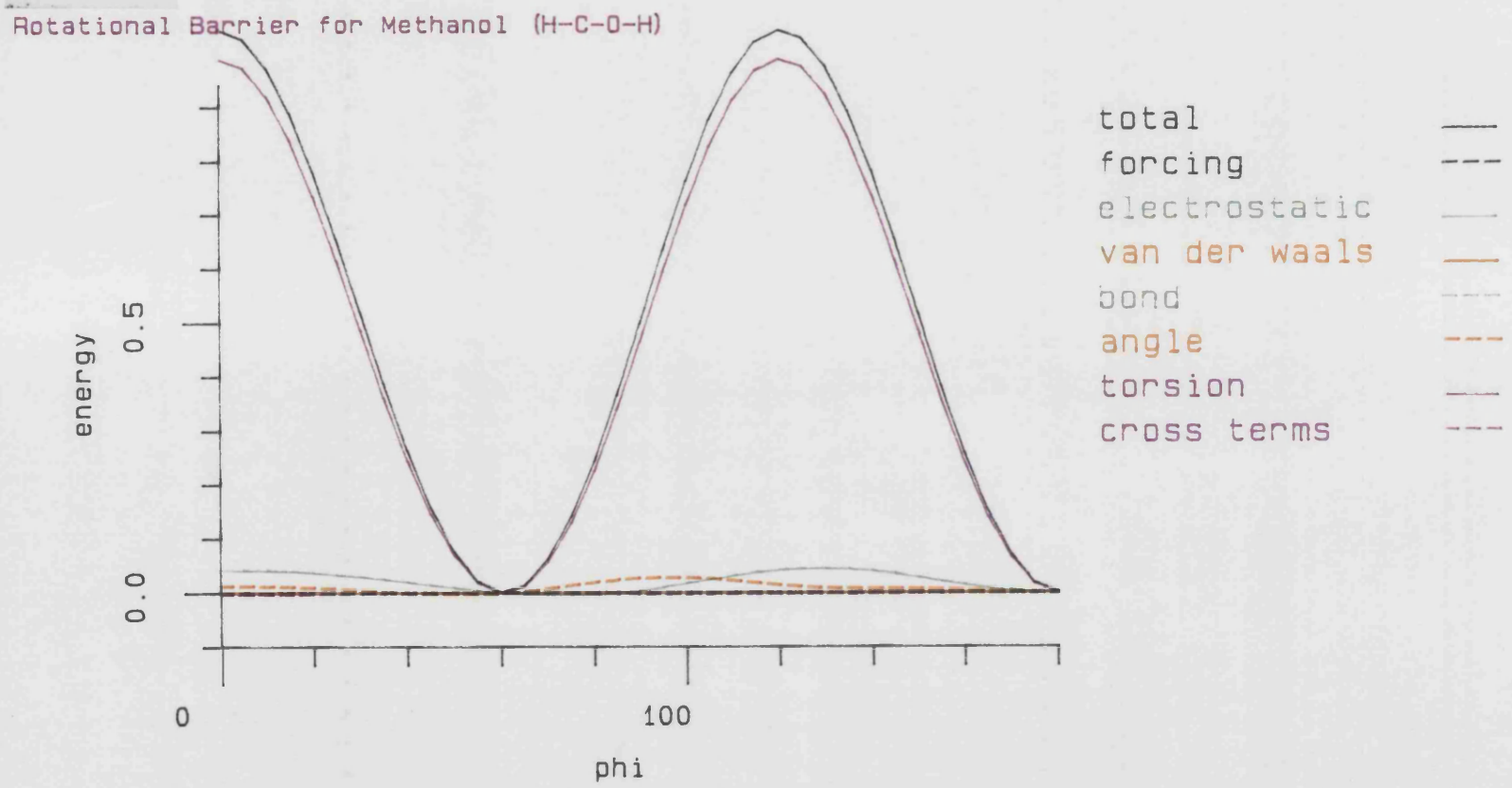

Rotational Barrier for Methanol (H-C-O-H)
energy
0.5
0.0
0
100
phi
total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

Figure 13a

Rotational Barrier for Ethanol (C-C-O-H)

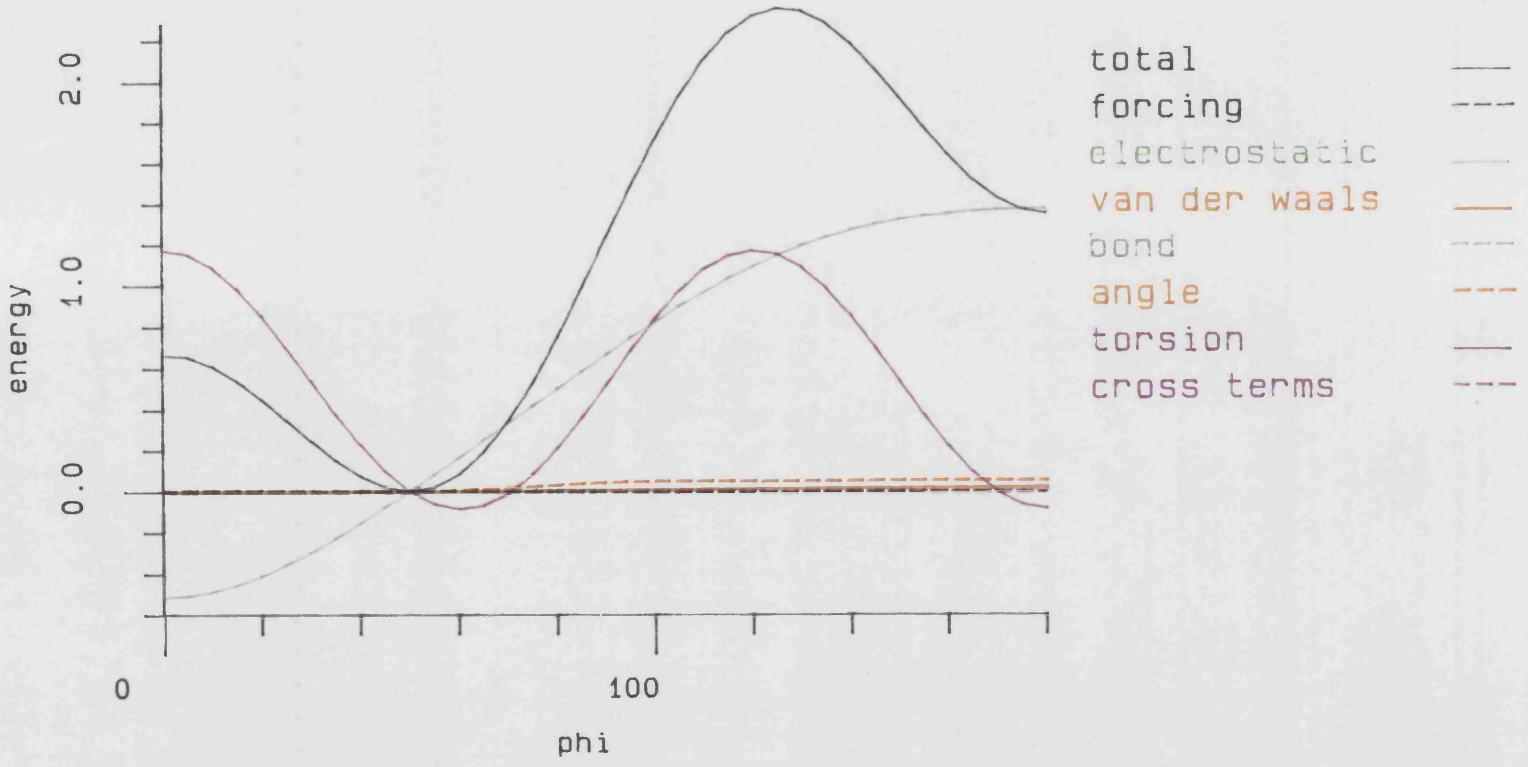


Figure 13b

Rotational Barrier for Ethanol (C-C-O-H) V1= 1.00

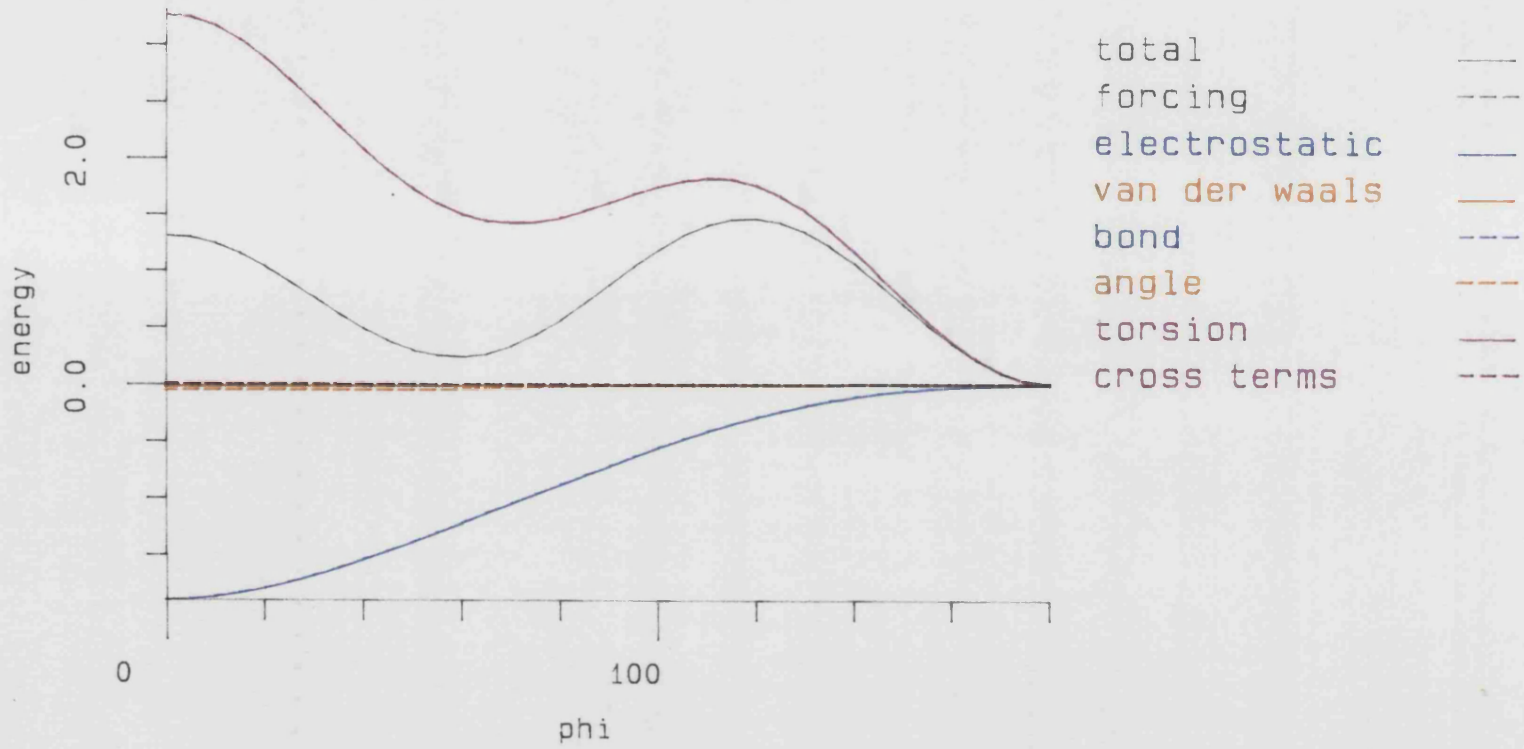


Figure 14a

Rotational Barrier for i-Propanol (H-C-O-H)

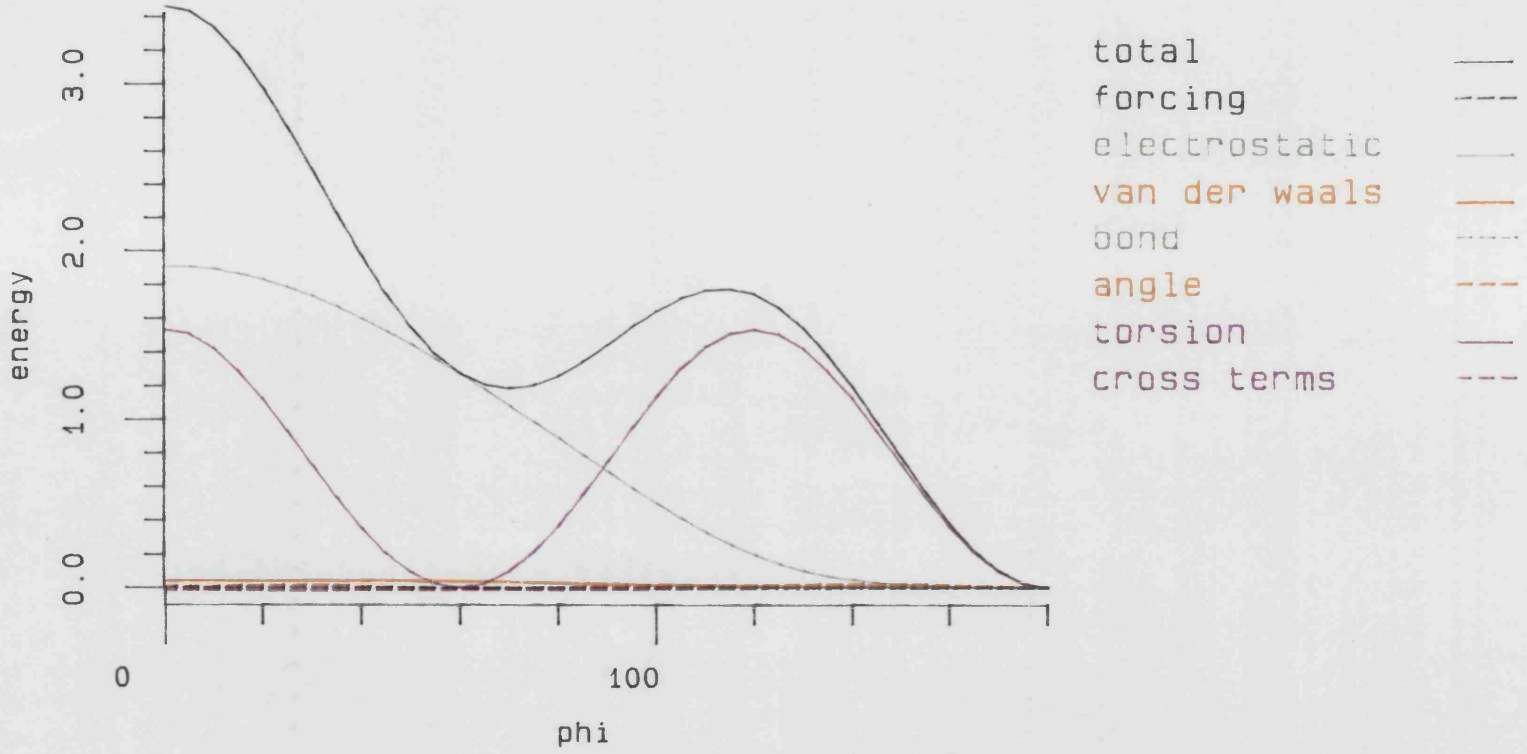


Figure 14b

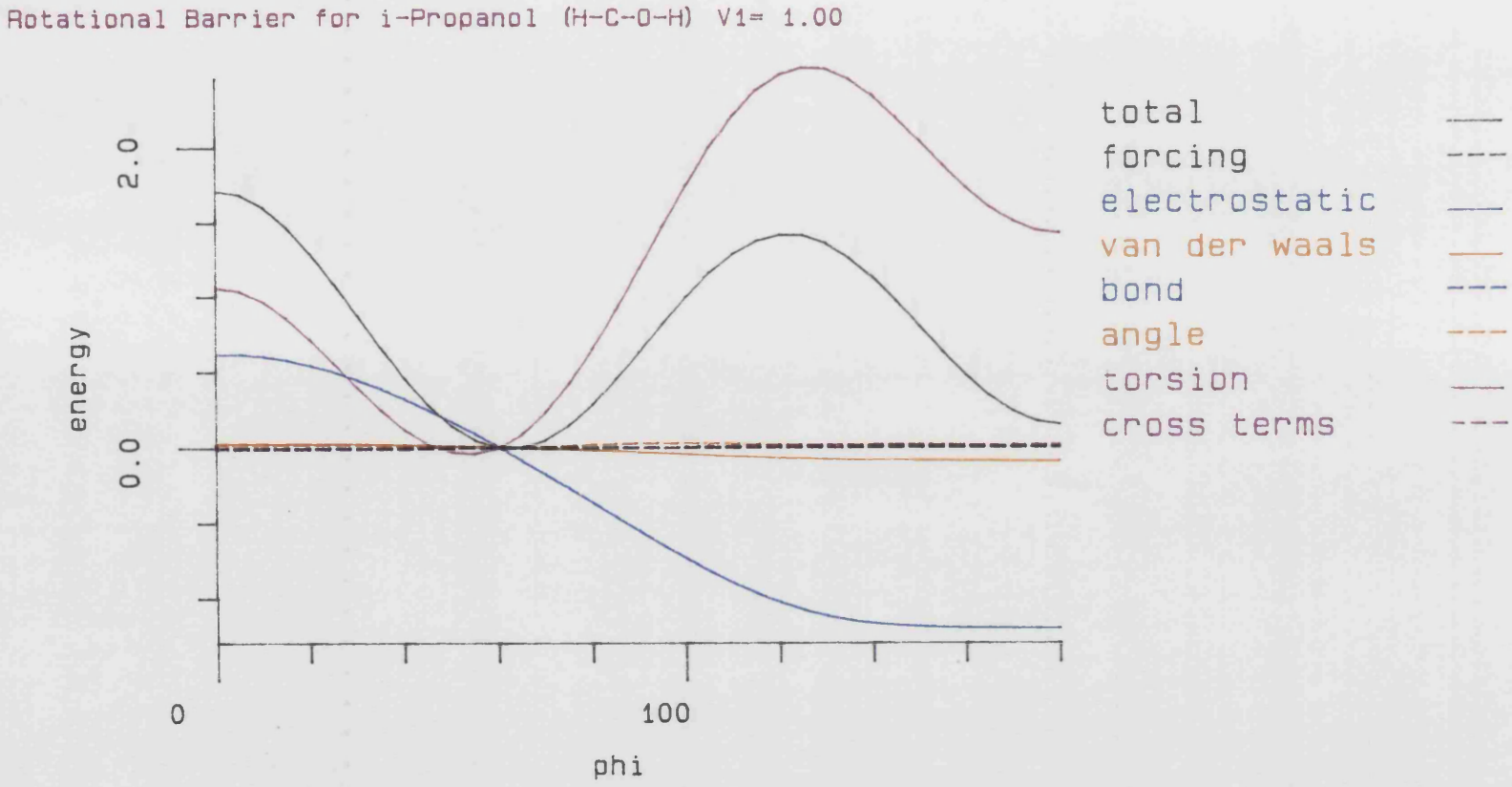
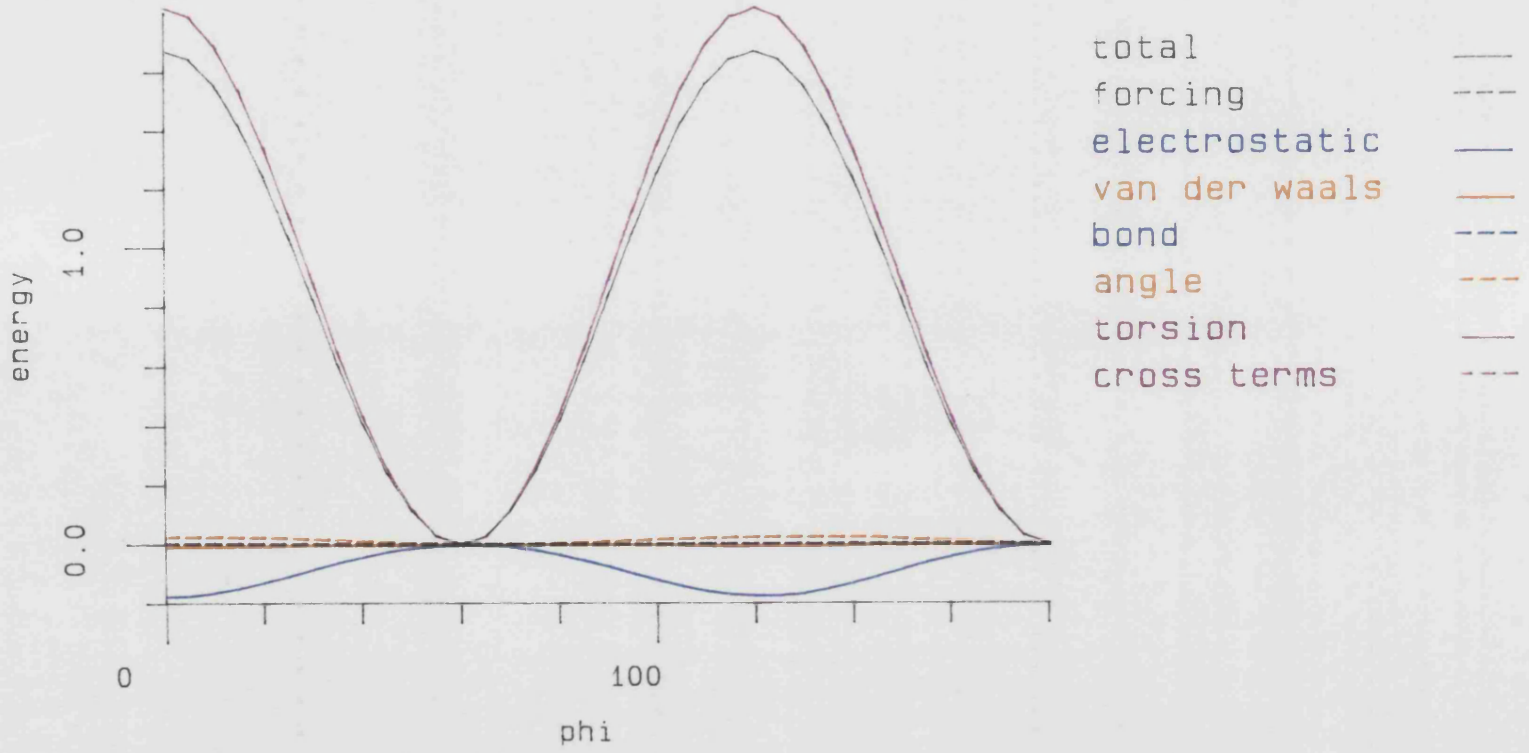

Rotational Barrier for i-Propanol (H-C-O-H) V1= 1.00
2.0
0.0
energy
0
100
phi
total
forcing
electrostatic
van der waals
bond
angle
torsion
cross terms

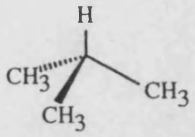
Figure 15

Rotational Barrier for t-Butanol (C-C-O-H)

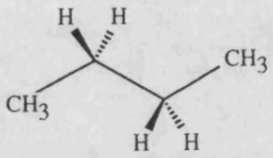


Appendix III

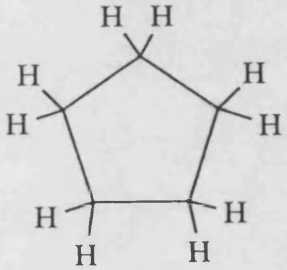
Model Compound Structures



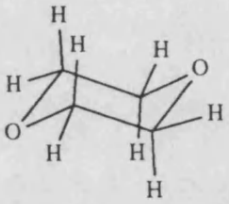
i-Butane



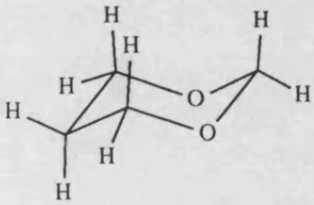
n-Butane



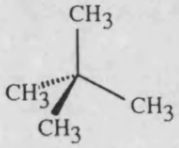
Cyclopentane



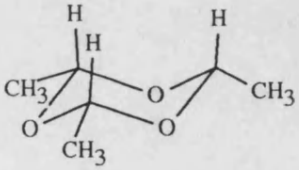
1,4-Dioxane



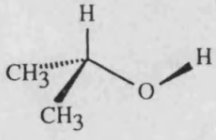
1,3-Dioxane



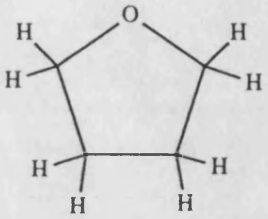
Neopentane



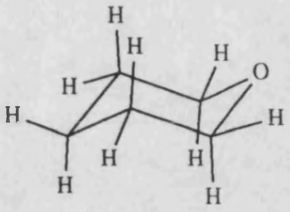
Paraldehyde



i-Propanol



Tetrahydrofuran



Tetrahydropyran